pects of diminished capacity with increasing age in those excessively dependent on external sources for self-esteem may cause the manic-depressive to strengthen his achievement values since he perceives them as a means of gaining needed approval. With respect to rigidity of attitudes, the manic-depressives' high authoritarian attitudes as measured by the F scale and TFI are unmodified by education. In the general population, authoritarian attitudes decrease as a function of educational level.

The results of the present study, in addition to indicating that the manic-depressive performs as a high value Achiever, served to verify empirically findings of Cohen et al., concerning a number of manic-depressive character traits. The reformulation of psychoanalytically derived descriptions of manic-depressive character within the framework of McClelland's experimentally derived concepts of achievement striving provided for a more objective and rigorous evaluation of these traits. The generality of the findings of Cohen et al., were extended by the present results to a somewhat different subculture of predominately white Southern Protestants. Furthermore, empirical verification of a number of behavior traits attributed to manic-depressives lends inferential support to their etiological formulation.

The findings of relatively specific adult achievement and conformity traits in manic-depressives suggests that hereditary-constitutional factors, are probably not sufficient in themselves to account for manic-depressive psychosis. These findings are, to a certain extent, compatible with speculations by traditional psychoanalytic writers (Fenichel, 1945) about the etiology of manic-depressive psychosis in terms of oral character structure. The conformity behavior of manic-depressives, which obtains the approval of others for them, is consistent with the oral character's intense dependency on others. Oral character explanations of manic-depressive psychosis, however, typically neglect the specification of socio-environmental factors. The findings of the present study support those of several recent inquiries (Finley & Wilson, 1951; Gibson, 1958; Wilson, 1951) which are in accord with the stress by Cohen and her collaborators on the importance of interpersonal experience in the etiology of manic-depressive character.

The results of the present study indicated that McClelland's Achievement concepts provided a useful framework within which to study pathological character. The utility of the value Achievement concept, however, was limited by its relatively undeveloped status. Greater research leverage could be achieved by translating manic-depressive character into a personality type whose characteristics, correlates, and genesis have been more comprehensively studied. The concept of the authoritarian personality, studied by a wide range of psychological techniques (Adorno, Frenkel-Brunswik, Levinson, & Sanford 1950; Titus & Hollander, 1957), seems to offer considerable research potential in this respect. In the present study, manic-depressives scored high on measures of authoritarianism and conventional attitudes. This finding suggests that such correlates of high authoritarianism as intolerance of ambiguity, social imperceptiveness, and various types of rigidity may be productively explored with manic-depressives.

SUMMARY

The present study investigated the achievement and conformity behavior of manic-depressives within the framework of McClelland's concepts of achievement striving. On the basis of similarities between McClelland's high value Achievers and manic-depressives as described by Cohen et al., specific hypotheses about the performance of manic-depressives on tasks relevant to value Achievement were formulated. Other tasks were included to test the possibility that the achievement strivings of manic-depressives may be similar to those of the high need Achiever. Twenty-four carefully selected manic-depressives were matched with thirty nonpsychiatric controls for age, education, and literacy level. All of the subjects were male white veterans between the ages of 25 and 65, with a minimum of sixth-grade reading ability; most were born and reared in the South. Nine experimental measures were used.

The manic-depressives scored significantly higher than the controls on value Achievement, the F scale, and the TFI scale. There were no differences in the scores of manic-depressives

and controls on a need Achievement measure or on either of two performance tasks. Thus, the manic-depressives performed as high value Achievers in that they showed a strong conscious valuation of achievement and highly authoritarian and conventional attitudes. The effects of age and educational level on these measures added further support to the performance of manic-depressives as value Achievers. On none of the measures did the manic-depressives perform as high need Achievers.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Arieti, S. The decline of manic-depressive psychosis: Its significance in the light of dynamic and social psychiatry. *The scientific papers of the 113th annual meeting of the Amer. Psychiat. Assn. in summary form*, No. 119, Wash., D. C.: Amer. Psychiat. Assn., 1957.

Atkinson, J. W., & Reitman, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

Becker, J. An analysis of manic-depressive character in terms of achievement and conformity. Unpublished doctoral dissertation, Duke Univer., 1958.

Brown, R. W. A determination of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.

Cohen, Mable B., Baker, Grace, Cohen, R. A., Fromm-Reichmann, Frieda, & Weigart, Edith V. An intensive study of twelve cases of manic-depressive psychosis. *Psychiatry*, 1954, **17**, 103–137.

DeCharms, R., Morrison, W. H., Reitman, W., & McClelland, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 414–423.

Fenichel, O. *The psychoanalytical theory of neuroses.* New York: Norton, 1945.

Finley, Cecile B., & Wilson, D. C. The relation of the family to manic-depressive psychosis. *Dis. nerv. System*, 1951, **12**, 39–43.

Foster, V. H., & Goddard, H. H. The Ohio-Literacy Test. *Pedagogical Seminary*, 1924, **31**, 340–351.

Gibson, R. W. The family background and early life experience of the manic-depressive patient. *Psychiatry*, 1958, **21**, 71–90.

Hildreth, H. M. A battery of feeling and attitude scales for clinical use. *J. Clin. Psychol.*, 1946, **2**, 214–221.

Kallman, F. J. *Heredity in health and mental disorder.* New York: Norton, 1953.

Levinson, D. J., & Huffman, Phyllis E. Traditional family ideology and its relation to personality. *J. Pers.*, 1955, **23**, 251–273.

Lewis, N. D. C., & Piotrowski, Z. A. Clinical diagnosis of manic-depressive psychosis. In P. H. Hoch & J. Zubin (Ed.), *Depression*. New York: Grune & Stratton, 1954.

McClelland, D. C. Some social consequences of achievement motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955*. Lincoln: Univer. Nebraska Press, 1955. Pp. 41–65.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

Riesman, D., Glazer, V., & Denney, R. *The lonely crowd*. New Haven: Yale Univer. Press, 1950.

Titus, H. E., & Hollander, E. P. The California F Scale in psychological research: 1950–1955. *Psychol. Bull.*, 1957, **54**, 47–65.

Wendt, H. W. Motivation, effort, and performance. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955. Pp. 448–459.

Williams, J. E. Mode of failure, interference tendencies, and achievement imagery. *J. abnorm. soc. Psychol.*, 1955, **51**, 573–580.

Wilson, D. C. Families of manic-depressives. *Dis. nerv. System*, 1951, **12**, 362–369.

Winterbottom, M. R. The relation of childhood training in independence to achievement motivation. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

(Received February 16, 1959)

Journal of Abnormal
and
Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University
URIE BRONFENBRENNER
Cornell University
ROGER BROWN
Massachusetts Institute of Technology
DONALD T. CAMPBELL
Northwestern University
LAUNOR F. CARTER
System Development Corporation
RICHARD CHRISTIE
Columbia University
MORTON DEUTSCH
Bell Telephone Laboratories
I. E. FARBER
State University of Iowa
DONALD W. FISKE
University of Chicago
N. L. GAGE
University of Illinois
ROGER W. HEYNS
University of Michigan
ROBERT R. HOLT
New York University
IRVING L. JANIS
Yale University
HAROLD E. JONES
University of California, Berkeley
HAROLD H. KELLEY
University of Minnesota
LEONARD S. KOGAN
Community Service Society of New York
DANIEL J. LEVINSON
Harvard Medical School
DAVID C. MCCLELLAND
Harvard University
LEO POSTMAN
University of California, Berkeley
ELIOT H. RODNICK
Duke University
NEVITT SANFORD
University of California, Berkeley
SEYMOUR SARASON
Yale University
RICHARD L. SOLOMON
Harvard University

Managing Editor: ARTHUR C. HOFFMAN
Promotion Manager: HELEN ORR

VOLUME 60
1960

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1960 by the American Psychological Association, Inc.

Bureau Edni. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated 8. 9. 70
Accs. No J-426

CONTENTS OF VOLUME 60

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

PAGE

Adams, J. S., & Hoffman, B. The Frequency of Self-Reference Statements as a Function of Generalized Reinforcement 384

Altman, I., & McGinnies, E. Interpersonal Perception and Communication in Discussion Groups of Varied Attidudinal Composition 390

Argyle, M. See Mohanna, A. I.

Atkinson, J. W., Bastian, J. R., Earl, R. W., & Litwin, G. H. The Achievement Motive, Goal Setting, and Probability Preferences 27

Atkinson, J. W., & Litwin, G. H. Achievement Motive and Test Anxiety Conceived as Motive to Approach Success and Motive to Avoid Failure 52

Barthol, R. P. See Goldstein, M. J.

Bastian, J. R. See Atkinson, J. W.

Becker, J. Achievement Related Characteristics of Manic-Depressives 334

Behringer, R. D. See Ziller, R. C.

Benedetti, D. T., & Hill, J. G. A Determiner of the Centrality of a Trait In Impression Formation 278

Benson, W. M. See Young, R. K.

Berkowitz, L. Repeated Frustrations and Expectations in Hostility Arousal 422

Berkowitz, L. Some Factors Affecting the Reduction of Overt Hostility 14

Berkowitz, L., & Cottingham, D. R. The Interest Value and Relevance of Fear Arousing Communications 37

Bieri, J. Parental Identification, Acceptance of Authority, and Within-Sex Differences in Cognitive Behavior 76

Borg, W. R. Prediction of Small Group Role Behavior from Personality Variables 112

Borgatta, E. F. The Stability of Interpersonal Judgments in Independent Situations 188

Brehm, J. W. Attitudinal Consequences of Commitment to Unpleasant Behavior 379

Broverman, D. M., Jordan, E. J., Jr., & Phillips, L. Achievement Motivation in Fantasy and Behavior 374

Carpenter, J. A. See McDonnell, G. J.

Cattell, R. B., & Scheier, I. H. Stimuli Related to Stress, Neuroticism, Excitation, and Anxiety Response Patterns: Illustrating a New Multivariate Experimental Design 195

Cavanaugh, D. K., Cohen, W., & Lang, P. J. The effect of "Social Censure" and "Social Approval" on the Psychomotor Performance of Schizophrenics 213

Chaney, Marilyn V., & Vinacke, W. E. Achievement and Nurturance in Triads Varying in Power Distribution 175

Chapman, L. J. Confusion of Figurative and Literal Usages of Words by Schizophrenics and Brain Damaged Patients 412

Chapman, L. J. See Usdansky, G.

Chatterjee, B. B., & Eriksen, C. W. Conditioning and Generalization of GSR as a Function of Awareness 396

Cline, V. B., & Richard, J. M., Jr. Accuracy of Interpersonal Perception—A General Trait? 1

Cohen, E., & Cohen, B. D. Verbal Reinforcement in Schizophrenia 443

Cohen, W. See Cavanaugh, D.

Consalvi, C. See Harvey, O. J.

Coopersmith, S. Self-Esteem and Need Achievement as Determinants of Selective Recall and Repetition 310

Cottingham, D. R. See Berkowitz, L.

Couch, A., & Keniston, K. Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable 151

Crow, W. R. See Fordyce, W. E.

PAGE

D'AMATO, M. R., & GUMENIK, W. E. Some Effects of Immediate Versus Randomly Delayed Shock on an Instrumental Response and Cognitive Processes.......... 64
DAVENPORT, G. See GILBERSTADT, H.
DAVIS, A. D. Some Physiological Correlates of Rorschach Body Image Productions.......... 432
DECHARMS, R. See ROSENBAUM, M. E.
DESOTO, C. B. Learning a Social Structure.......... 417
DIVESTA, F. J., & MERWIN, J. C. The Effects of Need-Oriented Communications on Attitude Change. 80
EARL, R. W. See ATKINSON, J. W.
EIGENBRODE, C. R., & SHIPMAN, W. G. The Body Image Barrier.......... 450
ERIKSEN, C. W. See CHATTERJEE, B. B.
FILLENBAUM, S. See LAMBERT, W. E.
FISHER, S. Head-Body Differentiations in Body Image and Skin Resistance Level.......... 283
FORDYCE, W. E., & CROW, W. R. Ego Disjunction: A Failure to Replicate Trehub's Results.......... 446
FULKERSON, S. C. Individual Differences in Reaction to Failure-Induced Stress.......... 136
GAGE, N. L., & CHATTERJEE, B. B. The Psychological Meaning of Acquiescence Set: Further Evidence. 280
GARDNER, R. C. See LAMBERT, W. E.
GILBERSTADT, H., & DAVENPORT, G. Some Relationships between GSR Conditioning and Judgments of Anxiety.......... 440
GOLDSTEIN, M. J., & BARTHOL, R. P. Fantasy Responses to Subliminal Stimuli.......... 22
GUMENIK, W. E. See D'AMATO, M. R.
HARVEY, O. J., & CONSALVI, C. Status and Conformity to Pressures in Informal Groups.......... 182
HILL, J. G. See BENEDETTI, D. T.
HODGSON, R. C. See LAMBERT, W. E.
HOFFMAN, B. See ADAMS, J. S.
HOLTZMAN, W. H. See YOUNG, R. K.
JENKINS, J. J., & RUSSELL, W. A. Systematic Changes in Word Association Norms: 1910–1952.......... 293
JORDAN, E. J., JR. See BROVERMAN, D. M.
KENISTON, K. See COUCH, A.
LAMBERT, W. E., HODGSON, R. C., GARDNER, R. C., & FILLENBAUM, S. Evaluational Reactions to Spoken Languages.......... 44
LANG, P. J. See CAVANAUGH, D.
LEACH, W. W. Nystagmus: An Integrative Neural Deficit in Schizophrenia.......... 305
LESHAN, L., & REZNIKOFF, M. A Psychological Factor Apparently Associated with Neoplastic Disease 439
LESTER, J. R. Production of Associative Sequences in Schizophrenia and Chronic Brain Syndrome. 225
LEWIN, M. H. See MARTIN, M.
LIPETZ, M. E. The Effects of Information on the Assessment of Attitudes by Authoritarians and Nonauthoritarians.......... 95
LITWIN, G. H. See ATKINSON, J. W.
LUNDY, R. M. See MARTIN, B.
LYONS, J. An Interview with a Mute Catatonic.......... 271
MCDONNELL, G. J., & CARPENTER, J. A. Manifest Anxiety and Prestimulus Conductance Levels.......... 437
MCGINNIES, E. See ALTMAN, I.
MCGUIRE, W. J. Cognitive Consistency and Attitude Change.......... 345
MCGUIRE, W. J. Direct and Indirect Persuasive Effects of Dissonance-Producing Messages.......... 354
MCILVAINE, MARY T. See WESSMAN, A. E.
MAHONE, C. H. Fear of Failure and Unrealistic Vocational Aspiration.......... 253
MANIS, M. The Interpretation of Opinion Statements as a Function of Recipient Attitude.......... 340
MARTIN, B., LUNDY, R. M., & LEWIN, M. H. Verbal and GSR Responses in Experimental Interviews as a Function of Three Degrees of "Therapist" Communication.......... 234
MASON, W. A. Socially Mediated Reduction in Emotional Responses of Young Rhesus Monkeys... 100

Page

Merwin, J. C. See DiVesta, F. J.

Milgram, N. A. Cognitive and Empathic Factors in Role-Taking by Schizophrenic and Brain Damaged Patients 219

Mogar, R. E. Three Versions of the F Scale and Performance on the Semantic Differential 262

Mohanna, A. I., & Argyle, M. A Cross-Cultural Study of Structured Groups with Unpopular Central Members 139

Moss, C. S. Brief Successful Psychotherapy of a Chronic Phobic Reaction 266

Palmer, S. Frustration, Aggression, and Murder 430

Pepitone, A., & Wilpizeski, C. Some Consequences of Experimental Rejection 359

Phillips, L. See Broverman, D. M.

Pine, F. Incidental Stimulation: A Study of Preconscious Transformations 68

Pisoni, S. See Salzinger, K.

Reitman, W. R. Motivational Induction and the Behavior Correlates of the Achievement and Affiliation Motives 8

Reznikoff, M. See LeShan, L.

Richards, J. M., Jr. See Cline, V. B.

Ricks, D. F. See Wessman, A. E.

Rogers, J. M. Operant Conditioning in a Quasi-Therapy Setting 247

Rosenbaum, M. E., & DeCharms, R. Direct and Vicarious Reduction of Hostility 105

Rosenberg, S. Cooperative Behavior in Dyads as a Function of Reinforcement Parameters 318

Ruebush, B. K. Interfering and Facilitating Effects of Test Anxiety 205

Russell, W. A. See Jenkins, J. J.

Rychlak, J. F. Recalled Dream Themes and Personality 140

Salzinger, K., & Pisoni, S. Reinforcement of Verbal Affect Responses of Normal Subjects during the Interview 127

Sapolsky, A. Effect of Interpersonal Relationships upon Verbal Conditioning 241

Scheier, I. H. See Cattell, R. B.

Shaw, M. E. A Note Concerning Homogeneity of Membership and Group Problem Solving 448

Shipman, W. G. See Eigenbrode, C. R.

Usdansky, G., & Chapman, L. J. Schizophrenic-Like Responses in Normal Subjects under Time Pressure 143

Vinacke, W. E. See Chaney, Marilyn V.

Wessman, A. E., Ricks, D. F., & McIlvaine, Mary T. Characteristics and Concomitants of Mood Fluctuation in College Women 117

Willis, R. H. Stimulus Pooling and Social Perception 365

Wilpizeski, C. See Pepitone, A.

Wilson, W. C. Extrinsic Religious Values and Prejudice 286

Worell, L. EEPS N Achievement and Verbal Paired-Associates Learning 147

Wurtz, K. R. Some Theory and Data Concerning the Attenuation of Aggression 134

Young, R. K., Benson, W. M., & Holtzman, W. H. Change in Attitudes toward the Negro in a Southern University 131

Ziller, R. C., & Behringer, R. D. Assimilation of the Knowledgeable Newcomer under Conditions of Group Success and Failure 288

Zimbardo, P. G. Involvement and Communication Discrepancy as Determinants of Opinion Conformity 86

Zuckerman, M. The Effects of Subliminal and Supraliminal Suggestion on Verbal Productivity 404

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

ACCURACY OF INTERPERSONAL PERCEPTION —A GENERAL TRAIT?[1]

VICTOR B. CLINE AND JAMES M. RICHARDS, JR.
University of Utah

When in Act I of *Pelleas and Melisande* one of the characters asks, "I have never for one instant seen clearly within myself; how then would you have me judge the deeds of others," a most crucial question is posed for those conducting research in the area of human judgment and interpersonal perception. One of its corollaries would be whether ability to perceive others accurately, if possible at all, is a general trait, or more specific and dependent on *whom* we are perceiving and what sort of judgment of them we are making. In the senior author's prior research in this area (Cline: 1955a; 1955b), the results suggested that ability to judge does have some generality. Later studies by Gage and Cronbach (1955), Crow and Hammond (1957), and Stone and Leavitt (1954), among others, suggested that on the contrary there was little relationship between accuracy of perception scores derived from two or more different instruments or procedures. The present research represents an attempt to determine more precisely if there is a general, global "trait" of perceiving others accurately that is susceptible to meaningful measurement.

METHOD

Subjects

The *S*s were 50 students, both male and female, in the 1958 summer session at the University of Utah, drawn from classes in psychology, education, sociology, and anthropology on a voluntary basis. Because of the small *N*, results were not analyzed by sex of judges. Such an analysis is planned in later studies. This group was more heterogeneous than is usual in university settings, since it included school teachers and others attending the summer session for special purposes. Although the *N* is rather small for a correlational study, it was thought to be sufficiently large for a study aimed at answering the question of generality.

Stimulus Materials

One of the more difficult problems in studying interpersonal perception is that of holding cues constant for all observers. In our prior research (Cline, 1955a; Cline & Richards, 1958), the most effective and useful procedure for doing this and at the same time providing the observers with a vivid impression of the "stimulus person" was found to be the use of sound motion pictures of an interview situation. Black and white film, employed in the initial stages of this research, proved to obscure too many cues relating to dress, complexion, etc. Therefore all interviews used in the present study were filmed in color.

These interviews as nearly as possible represented a cross section of adults in the Salt Lake City metropolitan area. The interviews were filmed at three different locations: an anteroom to a suburban supermarket, a downtown theater lobby, and a university research laboratory. Interviewees[2] were recruited from supermarket customers, off the street, and from university classes. All interviewees were given a $5.00 merchandise order for participating. All interviewees entered the situation "cold." The camera and sound equipment were in full view of the interviewee during the filming of the interview. Experience with this technique has indicated that their presence does not significantly affect the spontaneity of the interview.

All interviews were conducted by an actor, a member of the University of Utah theater staff. The interview procedure, while permitting the interviewer some freedom, was sufficiently structured to insure equivalence over interviews. The interview probed the following areas: (*a*) personal values, (*b*) personality strengths and weaknesses, (*c*) reaction to the interview, (*d*) hobbies and activities, (*e*) self-conception, and (*f*) temper. Each interviewee was given a brief explanation of the purpose of the research, followed by a "warm-up" period to gain rapport. The interviewer then gave a signal to the camera and sound man to begin filming, and started the interview proper.

Eventually 25 interviews were filmed in sound and color. From this larger group, 10 films were selected for the present research on the basis of technical excellence of sound and color, significant content of the interview, and a pilot study reported in greater detail elsewhere (Cline & Richards, 1958) which employed 20 films. Selection was made on the basis of the extent to which films discriminated good judges from poor judges, and thus closely approximates standard item analysis. Selection on this basis does, of course, to a certain extent "stack the cards" in favor of a general trait. This criticism can be made, however, of any internal consistency item analysis.

The 10 interviewees finally selected included: (*a*) a 19-year-old engaged male geology major at the University of Utah, (*b*) a 36-year-old married female, of Greek

---

[1] This study was supported by Contract NONR 1288 (04), Group Psychology Branch, Office of Naval Research.

[2] Throughout this paper "interviewee" refers to persons in the films about whom judgments are made.

parentage, with three children, (c) a 40-year-old married male occulist of Italian parentage, (d) a 22-year-old divorced female waitress with no children, (e) a 32-year-old married grammar school teacher with three children, (f) a 65-year-old widow with two married children working as supervisor of a church cafeteria, (g) a 22-year-old single male Mexican-American working in a meat packing plant, (h) a 17-year-old single female high school student of German parentage, (i) a 41-year-old married male IBM worker with one child, and (j) a 25-year-old divorced female with two children living with her parents. These interviewees are therefore quite diverse with regard to age, sex, and background. No analysis by sex of interviewee was made in the present study, but such analysis is intended in later studies using larger samples.

Judging Instruments

Following the filming of the interview, an intensive study was made of each of 24 interviewees, the 25th being unable to participate further. The following information was obtained in the course of several visits to the interviewee's home. It included an intensive interview on life history and personal habits and the pencil and paper tests listed below:

(a) MMPI
(b) Strong Vocational Interest Blank
(c) California Psychological Inventory
(d) a word association test
(e) Otis Intelligence Test
(f) a multiple choice sentence completion test (taken three times)
(g) the Gough Adjective Check List (taken three times)
(h) an 80-item cluster from the MMPI (taken three times)
(i) a 50-item Likert Trait Rating Scale (self-rating taken three times)

In addition, five close associates were given intensive interviews concerning their knowledge of the interviewee's behavior and history, and they rated the interviewee on the 50-item Trait Rating Scale. These interviews were conducted by a Ph.D. psychologist, a graduate student working toward his Ph.D. in psychology, and a second year law student. All interviews with a given interviewee and his associates were conducted by the same interviewer.

To select these associates, the interviewee was asked to nominate five people who knew him well and for a period of several years. An attempt was made to get both sexes, persons from different generations and status levels (parents, peers, boss, employees, etc.) These five people were then contacted directly by a member of the research team and the interview arranged.

From this information, the researchers constructed five judging instruments for judges to fill out after seeing each filmed interview. These were:

(a) *Behavior Postdiction.* On this judging instrument, the judge is required to "postdict" real life behavior on the part of the person seen in the filmed interview. A sample item would be:

8. When Mr. M gets angry he usually:
— (a) ends up cursing obscenely
— (b) gets so upset he has to walk out or leave
— (c) tends to keep it to himself and hides his feelings

Items were tailored to each individual separately, and were not common to interviews. The one correct alternative was determined on the basis of the life history interview with the interviewee and his five close associates. Only that behavior where there was agreement by all persons interviewed was used in constructing items. Considerable effort was made to insure that the incorrect alternatives were completely inaccurate and not even partially right. Items were so constructed that in the opinion of the authors a judge could not achieve a high score by always checking the most socially desirable alternative. There was, however, no formal control of social desirability on this or any of the following instruments, since the subjective judgment of the authors was used as the measure of social desirability throughout. Ten items were constructed for each interviewee, making a total of 100. Score on the Behavior Postdiction is the number of correct alternatives checked.

(b) *Trait Rating.* On this instrument, the judge rated the interviewee on a six-category Likert scale for each of 25 traits. The same 25 traits were used for all interviewees and included both favorable and unfavorable characteristics such as efficient, a leader, impulsive, and unstable. The criterion was obtained from the mean rating on the same trait made by the five close associates of the interviewee. The score on Trait Rating is the average of the squared discrepancies between the judge's ratings on individual traits and the criterion ratings. Cronbach (1955) has developed the rationale for this scoring system extensively. In contrast to the scoring system used for the other judging instruments in this research, this is an error score. In order to prevent confusion, in reporting correlations between Trait Rating and other instruments, the signs have been so arranged that a positive sign means that a good judge as measured by Trait Rating also tended to be a good judge as measured by the other instrument.

(c) *Opinion Prediction.* This instrument consisted of 20 MMPI items for each interviewee, such as, "I fall in love rather easily." The judge was required to check whether the interviewee answered each question True or False. Only those items that the interviewee answered consistently on each of the three times that the 80-item cluster from the MMPI was administered were included. As nearly as possible these items were selected so that 10 had been answered True and 10 had been answered False. Items were so selected that, in the opinion of the authors, judges could not obtain high accuracy scores on the basis of social desirability alone. Since this scale was constructed from each individual's interviewee's responses, the items included were not necessarily the same for any two different interviewees, although a few items might be common. There was a total of 200 items on this test, and the score is the total number correct.

(d) *Sentence Completion.* On this instrument, the task for the judge was determining how the interviewee had answered multiple-choice sentence completion items. Again only those items were included that the interviewee had answered in the same way on each of three separate occasions. On this instrument, as on the preceding ones, in so far as possible items were so chosen that judges could not in the opinion of the authors get a high accuracy score on the basis of social desirability of the answers. There are 100 items on this test, and the score is the total number correct. A sample item is:

4. When I make a mistake, I . . .
 — (*a*) am embarrassed
 — (*b*) laugh it off
 — (*c*) don't give a damn

Since items were selected on the basis of each individual interviewee's responses, the particular items included differed from one interviewee to another.

(*e*) *Adjective Check*. This instrument required the judge to determine which one of a pair of adjectives an interviewee had checked as being descriptive of himself. A sample item is:

18. — (*a*) Resourceful
 — (*b*) Cheerful

In each pair the correct choice was an adjective that the interviewee had checked as being descriptive of himself on each of three separate occasions, and the incorrect choice was an adjective that the interviewee had failed to check as being descriptive of himself on these three occasions. In the judgment of the researchers, the adjectives in each pair were roughly equivalent as to social desirability, preventing the use of this variable in obtaining a high accuracy score. There were 20 such pairs for each interviewee, a total of 200 items in all, with the particular pairs included differing from one interviewee to another. The score is the total number correct.

These instruments as reported represent final forms. Several of the instruments have been extensively revised and refined on the basis of previous pilot studies. A more detailed report of these revisions is available elsewhere (Cline & Richards, 1958).

Procedure

All judges in this research made their judgments in groups. At the start of each judging session, the judges filled out forms dealing with necessary background information. They were then given a brief period for inspecting the instruments they would have to fill out. At the conclusion of this period, the film of the first interview was shown. The *S*s then filled out the judging instruments with regard to the first interviewee. The film of the second interview was then shown, and the judging instruments filled out with regard to the second interviewee. This procedure was repeated until all 10 films had been shown and all judging instruments completed. All judges saw the films in the same order. Approximately two hours were required for judges to view all 10 films and complete all judging instruments.

RESULTS

In addition to the scores on the various judging instruments a Total Judging score was obtained. This was computed by converting scores on each judging instrument to standard scores ($M = 50$, $\sigma = 10$) to insure equal weighting and adding across tests. In converting to standard scores, Trait Rating was converted from an error score to an accuracy score.

TABLE 1
RELIABILITIES OF JUDGING INSTRUMENTS

Trait Rating total	.72
Behavior Postdiction	.66
Sentence Completion	.67
Opinion Prediction	.67
Adjective Check	.79
Total Judging	.71

A first question relevant to the problem of generality of judging instruments is that of generality within judging instruments, which is to say their reliability. Here the question is one of whether, for a single judging instrument, a judge who judges the first five films well also does well with the second five films. The various reliability coefficients are presented in Table 1. The reliabilities for the Trait Rating and Total Judging were computed from the Spearman-Brown formula based on variances (Gulliksen, 1950) and a split between the first five films and the second five films. All other reliabilities were computed by using the Kuder-Richardson Formula 20.

There are several important considerations in interpreting these reliability coefficients. Guilliksen (1950) points out that the Kuder-Richardson reliability, when computed from Formula 20, is a lower bound to the reliability. In computing the Trait Rating, a simple first half-second half split was used, rather than an odd-even split. In computing the reliability of Total Judging, tests were combined with unit weight, no attempt being made to weight tests in such a way as to maximize the reliability. Therefore the reliabilities in Table 1 are, in general, lower than the maximum reliabilities for the various instruments.

A second question relevant to the problem of generality of judging ability is that of generality across judging instruments. This has to do with whether or not a good judge on one instrument also tends to be a good judge on another instrument. The intercorrelations, means, and standard deviations of the various judging measures are presented in Table 2. In

TABLE 2

INTERCORRELATIONS AMONG JUDGING MEASURES

($N = 50$)

Measure	Trait Rating Total	Behavior Postdiction	Sentence Compl.	Opinion Pred.	Adj. Check	Total Judging	M	σ
Trait Rating total	—						2.4273	.8139
Behavior Postdiction	.30	—					54.37	7.62
Sentence Completion	.48	.47	—				46.72	7.38
Opinion Prediction	.52	.50	.47	—			128.90	10.50
Adjective Check	.65	.24	.58	.54	—		156.90	11.76
Total Judging	.63	.44	.63	.65	.66	—	250.64	38.06

$r\ .05 = .27$ $r\ .01 = .35$

this table signs involving Trait Rating have been so arranged that a positive correlation means that a good judge on Trait Rating also tended to be a good judge on the other judging measure. The mean and standard deviation for Trait Rating are, however, *error* scores.

All correlations in Table 2 between Total Judging and the various individual judging instruments involve the total judging score with the contribution of the particular instrument under consideration eliminated.

Of the correlation coefficients in Table 2, only one would occur by chance with $p > .05$ and only two with $p > .01$. It therefore seems highly improbable that Table 2 represents chance correlations.

Cronbach (1955) has strongly and effectively criticized the use of global measures of accuracy of interpersonal perception. In the present research, the scoring system for Trait Rating was specifically chosen to permit the more analytical study of accuracy scores suggested by Cronbach. (At the present time Cronbach advocates still another procedure for studying interpersonal perception, even going so far as to suggest that accuracy scores of the type used in this research be discarded altogether. The present authors feel, however, that the type of analysis worked out by Cronbach earlier is the more appropriate for their purposes.) When this procedure is followed, the scores on Trait Rating are analyzed into four components. These four components can be further analyzed into variances and correlation terms, but the finer analysis was not attempted here since the four basic components in difference score form appeared more immediately relevant. (The authors intend to make the further analysis in later studies.) These four components are:

1. *Elevation.* This component represents the extent to which the grand mean of an individual judge, taken over all 10 films and all 25 traits, differs from the criterion grand mean. Cronbach is of the opinion that this component has to do with the way in which a judge uses the rating scale, and has little to do with what is usually meant by "accuracy of interpersonal perception."
2. *Differential Elevation.* This component has to do with the judge's ability to predict the deviation from the grand mean of the mean of individual *interviewees* taken over all traits. It therefore measures the variance of an individual judge's ratings. Cronbach states that in some cases this can be meaningful. In the present study, however, it too seems related mainly to the way in which the judges use the rating scale. It should therefore have little to do with accuracy of perception.
3. *Stereotype Accuracy.* This measures the ability of the judges to predict the deviation from the grand mean of the mean of individual traits taken over all interviewees. This should be an important part of accurate interpersonal perception. It measures what Bronfenbrenner, Harding, and Gallwey (1958) have referred to as Sensitivity to the Generalized Other.
4. *Differential Accuracy.* This component measures the ability of judges to predict differences between judges on each trait considered separately. Cronbach states that this component represents an average over traits. It has been referred to by Bronfenbrenner and his associates (1958) as Interpersonal Sensitivity, and should also be important in interpersonal perception. The actual computational procedure used in deriving this component was the subtraction of the other three components from the Trait Rating total score.

The correlations of these components, together with their means and standard deviations, are presented in Table 3. The data clearly confirm Cronbach's recommendation that global accuracy of perception scores be analyzed.

In Table 3 the correlations between the

TABLE 3

Correlations Between Trait Rating Components and Other Judging Measures

$N = 50$

Measures	Elevation	Differential Elevation	Stereotype Accuracy	Differential Accuracy	Trait Rating Total Score
Elevation	—				
Differential Elevation	.12	—			
Stereotype Accuracy	.04	.01	—		
Differential Accuracy	.29	.29	.16	—	
Trait Rating Total Score	.25	.24	.16	.25	—
Behavior Postdiction	−.05	−.23	.66	.01	.30
Sentence Completion	−.20	.08	.60	.25	.48
Opinion Prediction	.20	.00	.59	.28	.52
Adjective Check	.05	.23	.44	.56	.65
Total Judging	.16	.10	.76	.53	.63
M	.1015	.1162	.7704	1.4392	2.4273
σ	.1154	.0589	.3911	.5759	.8139

$r_{.05} = .27$ $r_{.01} = .35$

various components and the Trait Rating Total Score involve the total score with the contribution of the particular component under consideration eliminated. In computing the correlations between these component scores and the Total Judging score, however, it proved to be impossible to eliminate the contribution of the particular component under consideration, since the components represent the relative contribution to the variance of the error scores, a peculiarity of the Trait Rating scoring system. No technique was known to derive the relative contribution of these components to the variance of accuracy scores.

In Table 3 the signs of correlations involving Trait Rating and its components have again been arranged so that a positive correlation means what it would ordinarily be taken to mean. The means and standard deviations are, however, of error scores.

No reliabilities are available for these components because the amount of calculation involving in obtaining them is prohibitive. At the present time, however, a program for a high speed computer is being written to obtain these reliabilities.

Discussion

The reliabilities in Table 1 and intercorrelations in Table 2 tend strongly to indicate that a general, global ability to judge others accurately can be meaningfully measured. The data in Table 3 indicate, however, that this global ability involves a complicated "dynamic" process, or, more precisely, a factorially complex one, and helps to clarify what it is that is general in the global measure. Stereotype Accuracy accounts for a large portion of the generality. However, even after this component is eliminated, considerable generality remains, which seems to be related mainly to the Differential Accuracy component. These two components are relatively independent. It thus appears that a good judge may be accurate because he has an accurate stereotype or because he is able to predict specific differences between individuals (or both).

Bronfenbrenner and associates (1958) have recently come to identical conclusions concerning the two independent components in accuracy of interpersonal perception. Since his research used quite different procedures, it and the present study offer striking confirmation of each other.

The results of this study, in other words, indicate that there is a general ability to perceive others accurately. This general ability, however, consists of two (at least) independent parts: Sensitivity to the Generalized Other and Interpersonal Sensitivity in Bronfenbrenner's terminology, or Stereotype Accuracy and Differential Accuracy in Cronbach's terminology.

It may appear paradoxical (or inappropriate) to conclude that at the same time the ability to perceive others accurately is general and that it consists of two independent components. The authors are of the opinion, however, that these seemingly conflicting conclusions mean that the ability to perceive others accurately is factorially complex and that the independent components reflect this complexity. It is therefore still meaningful to talk about a general trait. There are numerous precedents for this in psychology. Perhaps the clearest example is the variable Intelligence, which has been demonstrated to consist of several independent components but is still treated as a meaningful General trait. In investigations of intelligence, even those studies indicating a G factor have typically also shown several independent group factors.

Since the conclusions of this study differ considerably from those of Crow and

Hammond (1957), it is important to discuss methodological differences that might account for these different conclusions. Crow and Hammond used a wider variety of tasks, including some, such as sociometric measures, which seem only remotely related to what would usually be meant by "accuracy of interpersonal perception." Crow and Hammond also used a different scoring procedure, one that is difficult to evaluate. They are of the opinion that it results in a measure of the Differential Accuracy component. The results of this study indicate that Stereotype Accuracy should be included in studies of interpersonal perception, especially in seeking an answer to the question of generality. In those cases in which Crow and Hammond do deal with Stereotype Accuracy, their results also tend to indicate that this component is both reliable and general. Finally, in the present study the measures went through several refinements through procedures closely approximating standard item analysis. This apparently was not done in the Crow and Hammond study. In view of these differences, the authors hold that insofar as Crow and Hammond's study represents a criticism of naive operationism in the study of interpersonal perception, their point is well taken that the *assumption* of a general judging ability over diverse tasks is inappropriate in the absence of supporting empirical data. In view of the present demonstration of considerable generality over heterogeneous items and heterogeneous persons to be judged, however, Crow and Hammond's results cannot be considered to disprove a general ability to perceive people accurately.

One possible criticism of the conclusions of the present research could be derived from the methodological work of Gage, Leavitt, and Stone (1956). These authors hold that any generality of judging ability is likely to be due to real similarity between the persons judged. This view would suggest that selection of films on the basis of their discrimination of good judges from poor, as was done in this research, would result in selection of films in which the interviewees are more similar to one another than they are to the interviewees in the eliminated films. In order to check this possibility, a special key was prepared for the Trait Rating, based on the mean rating on each trait of the criterion scores of the 10 films included in the study. Then an "Index of Similarity" was computed for each of the 10 best films included and for each of the 10 films eliminated on the basis of the earlier pilot study by scoring the *criterion* ratings for each film against this special key. The two groups of films were then compared with regard to this Index. The results of this comparison ($t = 1.694$, $df = 18$) fail to support a significant difference between the groups of films in this respect. The small "chance" difference that does occur, however, is in the direction predicted by Gage, Leavitt, and Stone (1956). Generality of judging ability to the extent found in this study cannot, therefore, be explained on the basis of real similarity to one another of the persons judged.

SUMMARY

Recent investigations have suggested that there is little, if any, generality of ability to judge others accurately. This study indicated that on the contrary it is possible to obtain a measure of the ability to judge others accurately that is reliable and general over a variety of judging instruments. The ability to judge others consists of at least two relatively independent components, *Sensitivity to the Generalized Other* and *Interpersonal Sensitivity*. These components are interpreted as reflecting the factorial complexity of the over-all judging scores. The correlations between the different judging instruments cannot be explained on the basis of the real similarity to one another of persons judged.

REFERENCES

BRONFENBRENNER, U., HARDING, J., & GALLWEY, MARY. The measurement of skill in social perception. In D. C. McClelland (Ed.), *Talent and society*. New York: Van Nostrand, 1958.

CLINE, V. B. An analysis of the performance of 100 Air Force officers on the Social Acuity Test. USAF Pers. Train. Res. Cent. tech. Memo., 1955, Proj. No. 7730, No. AF 18(600)-8, Maxwell AFB. (a)

CLINE, V. B. Ability to judge personality assessed with a stress interview and sound-film technique. *J. abnorm. soc. Psychol.*, 1955, **50**, 183–187. (b)

CLINE, V. B., & RICHARDS, J. M., JR. Variables related to accuracy in interpersonal perception. *USN Ann. Rep.*, 1958, ONR 171-146, Dept. Psychol., Univer. of Utah.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

CROW, W. J., & HAMMOND, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **54**, 384–390.

GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

GAGE, N. L., Leavitt, G. S., & STONE, G. C. The intermediary key in the analysis of interpersonal perception. *Psychol. Bull.*, 1956, **53**, 258–266.

GUILLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.

STONE, G. C., & LEAVITT, G. S. *Generality of accuracy in perceiving standard persons*. Bur. of Educ. Res., Coll. of Educ., Univer. of Illinois, 1954. (Mimeo.)

(Received January 9, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

MOTIVATIONAL INDUCTION AND THE BEHAVIOR CORRELATES OF THE ACHIEVEMENT AND AFFILIATION MOTIVES[1]

WALTER R. REITMAN[2]
University of Michigan

McClelland, Atkinson, and their collaborators have developed content analytic systems which, applied to modified Thematic Apperception Test (TAT) protocols, yield scores interpreted as indices of motive strength. Measures of the achievement and affiliation motives obtained in this way have been employed in numerous investigations of memory, perception, and performance as functions of motive strength. Occasional inconsistent findings (cf. Reitman, 1957) and failures to replicate previously obtained results (Birney, 1956)[3] however, suggest a need for further analysis and more precise specification of the basic assessment and validation procedures.

Atkinson and Reitman (1956), investigating Atkinson's conception of aroused motivation as a function of latent motive strength and of environmentally determined expectancies that action will lead to attainment of a motive's goal, predicted and found performance at an arithmetic task to be significantly better in a high n Achievement group than in a low group under an Achievement-Orientation condition designed to maximize expectancies of attaining an achievement goal and to minimize all others. In a related investigation, Reitman and Atkinson (1958) found predictions of arithmetic task performance based on motive scores derived from stories written to the first four pictures of an eight-picture measure to be significantly better than those from stories for the last four pictures, a result they considered to have critical implications for experimental and clinical use of the TAT.

The present study derives from these two previous investigations and affords a further opportunity to evaluate the effects of picture cues, picture order, individual vs. group administration of performance measures, and and achievement, neutral or relaxed-affiliation instructions upon correlations between achievement or affiliation motive scores and performance. Three predictions follow from the conception used by Atkinson and Reitman (1956): (*a*) Significant correlations between motive indices and performance should be positive under all conditions. (*b*) Performance should correlate most highly with indices of n Achievement under Achievement-Orientation conditions. (*c*) Performance should correlate most highly with indices of n Affiilation under the Relaxed-Affiliation condition. A fourth prediction is suggested by Reitman and Atkinson's (1958) results. (*d*) For all experimental conditions under which n Achievement or n Affiliation is significantly related to performance, the motive index based on the first three pictures of the present six-picture measure should yield larger correlations than that based on the last three pictures.

METHOD

All *S*s participated in two separate experimental sessions. At the first session, one of two forms of the thematic apperceptive motive measure was administered. At the second session, *S*s worked at three performance tasks under one of five performance conditions. These five conditions, defined by a combination of group or individual administration of the performance measures with either achievement-orienting, neutral, or relaxed-affiliation instructions, were: (*a*) group, Achievement-Oriented—46 *S*s; (*b*) individual, Achievement-Oriented—53 *S*s; (*c*) group, Neutral—46 *S*s; (*d*) individual, Neutral—50 *S*s; (*e*) group, Relaxed-Affiliation—63 *S*s.

Subjects. A total of 258 males from five sections of the elementary psychology course at the University of Michigan participated in the two parts of this experiment during the fall semester, 1955–56, in fulfillment of the experimental participation requirement of the course.

---

[1] This report is based on a dissertation submitted to the Department of Psychology, University of Michigan, in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to the National Science Foundation and the Social Science Research Council for fellowship support while conducting the investigation, and to Daniel Katz, John W. Atkinson, David Birch, William Hays, and E. Lowell Kelly for invaluable aid and advice throughout the course of the Study.

[2] Now at the Graduate School of Industrial Administration, Carnegie Institute of Technology.

[3] Personal communications from R. N. Haber, April 1955 and April 1956.

Measurement of n Achievement and n Affiliation. All Ss wrote six four-minute thematic apperceptive stories under neutral classroom conditions 10 to 11 weeks before the administration of the performance measures. Two forms were used. The "Ach-pictures first" form consisted of three pictures frequently used to measure the achievement motive followed by three used to measure the affiliation motive. The two three-picture subsets were reversed to make up the second, "Aff-pictures first" form. Picture order within each subset was constant. The three achievement-related pictures used are those designated 1B, 1H, and 1D in McClelland, Atkinson, Clark, and Lowell (1953, p. 375). The affiliation-related pictures used are designated 3E, 3A, and 3B by Atkinson, Heyns, and Veroff (1954, p. 405). Pictures 1B, 1H, 3A, and 3B also were used in the two antecedent studies by Atkinson and Reitman (1956) and Reitman and Atkinson (1958). The first form was administered in three lecture sections (C, S, T), the second in two (A, F). Stories to all six pictures were scored for both the achievement and affiliation motives by experienced scorers, all of whom had obtained scoring correlations above .90 with one of the authors (Atkinson) of the scoring systems in checks of their scoring reliability.

Assessment of performance. To control for interaction between motivational induction instructions and aptitude, Quantitative (Q) and Linguistic (L) Aptitude scores from the ACE examination were obtained for 195 of the 258 Ss taking part in the experiment; the remaining 63 Ss were assigned to the Relaxed-Affiliation condition.

Each S used a booklet containing one of the three induction instructions and all three performance measures. Each induction instruction consisted of a longer introductory part, forming the first page of the booklet, and briefer "explanatory remarks" about each performance measure. The latter appeared together with the instruction for doing the task on a single page, with the intent of maintaining the motivational orientation induced by the introductory instruction.

The achievement-orienting instruction, stressing the sensitivity of the performance "tests" to care and effort and emphasizing that performance would be evaluated to see "how well you can really do," was designed to arouse a strong expectancy that excellent performance would lead to attainment of an achievement goal, and to minimize all other expectancies. The neutral instruction described the tasks as part of a routine demonstration intended to familiarize students with some typical psychological measures. The relaxed-affiliation instruction, in effect denying the possibility of evaluation by describing the tasks as "preliminary" and "unstandardized," and emphasizing instead E's appreciation of S's helping him "work the kinks out," was designed to arouse an expectancy that performance would be related to attainment of an affiliation goal and to minimize all other expectancies.

The first of the three performance measures was a set of three-step arithmetic problems used in the two antecedent studies. In each problem, two sets of three numbers each are added and subtracted separately; the final solution involves addition or subtraction of the two intermediate results. A spatial perception measure, taken from Thurstone and Thurstone (1943), was next. In each group of items, S must discriminate test figures which are rotations of an inspection figure from those which are not. A third measure consisted of items from a Number Matching Test developed by Sperber (1955), in which S inspects both members of a pair of numbers and indicates whether they are identical.

Immediately before the first performance measure, Ss read a short story and were told that they would have to recall it later. This was done in connection with another experiment, and reproduces an experimental procedure common to the two antecedent studies. Two other brief measures and a subsequent questionnaire introduced for other purposes are not analyzed here, except for one questionnaire item checking a hypothesis offered by Atkinson and Reitman (1956, p. 363) about the extent to which recall of the short story may have confounded motive-performance correlations anticipated in their study.

In the individual conditions, Ss were brought into one of eight small laboratory rooms on either side of a central corridor, one S per room. The rooms were closed, E conducting the experiment from the corridor. In both individual and group conditions, when all Ss had arrived, E instructed them to turn over the booklets which lay face down before them, read the instructions, and fill in any information requested. To minimize variance due to interaction between E and Ss, E did not otherwise communicate with Ss except to direct them to a new instruction or measure. The Ss worked nine, four, and three minutes on the arithmetic, spatial, and matching problems respectively. Except for the three Relaxed-Affiliation groups of approximately 21 Ss each, the average N per session was about seven.

RESULTS AND DISCUSSION

Performance as a function of achievement-orienting vs. neutral instructions and individual vs. group administration. The effectiveness of the printed achievement-orienting instructions is clear from a comparison of performance under the two achievement-oriented conditions with performance under the two neutral conditions. Means and *SD*s of correct performance scores for all five conditions are given in Table 1. Table 2 summarizes the three analyses of variance evaluating performance on each measure as a function of achievement-orienting vs. neutral instructions and individual vs. group administration of the response measures. The printed achievement-orienting as compared with neutral instructions yield a significant increment in number of correct solutions on all three measures ($P < .05$).

TABLE 1

MEAN CORRECT ON RESPONSE MEASURES AS A FUNCTION OF INSTRUCTION AND ADMINISTRATION

Performance Condition		N	Response Measure					
			Arithmetic		Spatial		Matching	
Instr.	Admin.		Mean	*SD*	Mean	*SD*	Mean	*SD*
Ach.	Group	46	42.85	12.46	50.83	22.83	47.35	10.30
Ach.	Indiv.	53	43.87	14.72	56.76	27.51	48.34	11.38
Neutr.	Group	46	36.91	13.04	46.20	19.63	43.70	9.59
Neutr.	Indiv.	50[a]	41.46	13.62	45.70	21.80	45.72	10.44
R.-Af.	Group	63	36.81	14.05	47.52	19.33	45.53	6.90

[a] $N = 48$ for spatial measure; two *S*s in the neutral instruction, individual administration condition did not attempt the spatial perception measure.

TABLE 2

ANALYSES OF VARIANCE OF CORRECT SCORES AS A FUNCTION OF INSTRUCTION AND ADMINISTRATION

Source	*df*	Response Measure					
		Arithmetic		Spatial		Matching	
		MS	*F*	*MS*	*F*	*MS*	*F*
Instr.	1	810.09	4.33*	3223.44	5.82*	475.68	4.24*
Admin.	1	381.90	2.04	363.42	.66	111.93	1.00
I × A	1	151.44	.81	509.49	.92	15.51	.14
Error	191[a]	187.02		553.53		112.17	

*$P = .05$.

[a] $df = 189$ for spatial measure; cf. fn. 5.

Effects of quantitative and linguistic aptitude on performance. As the previous analyses show no evidence of a significant difference between the individual and group conditions, or of interaction between induction instructions and individual vs. group administration, the two administration conditions were combined for analyses of the effects of ACE *Q* and *L* scores. To determine the effects of differences in aptitude, *S*s were divided into three roughly equal groups, first by *Q* and then by *L* scores. Analyses of variance evaluating performance on the three measures as a function of achievement-orienting vs. neutral instructions and of low, moderate, or high *Q* scores show quantitative aptitude to be a highly significant determinant of correct solutions on the arithmetic and spatial measures ($P < .01$) as well as on the matching measure ($P < .05$). Corresponding analyses provide no evidence that linguistic aptitude is related to performance on any of the measures. Finally, neither set of analyses of variance indicates that performance is significantly affected by interaction between induction instructions and either *Q* or *L* scores.

Motive indices and performance. The first prediction states that significant correlations between motive indices and performance should be positive under all experimental conditions. The significance of the motive-performance correlations presented in Table 3 as functions of the several parameters of this study therefore is evaluated with reference to one-tailed significance points.

According to the second prediction, the n Achievement index should correlate maximally with performance under Achievement-Orientation conditions. The tabled correlations do not support this prediction. Of the 12 *r*s between achievement motive scores and the number of items correct on each of the three performance measures under achievement-oriented conditions, only the one with the arithmetic measure for *S*s from Sections A and F working under the individual condition is positive and significant at the $P = .05$ point; 5 of the 12 *r*s are negative.

According to the third prediction, the n Affiliation index should correlate maximally with performance under the Relaxed-Affiliation condition. Of the six correlations between number of items correct on each of the three performance measures and n Affiliation scores based on all six pictures, only one is positive and significant ($P = .05$); four of the six are negative. These results provide no support for the prediction, particularly insofar as a set of larger correlations (up to. 62) between n Affiliation and performance occur under the achievement-oriented, group administration condition, and this is the condition in which such a relationship is least likely to occur according to the theoretical conception that generates the prediction.

The fourth prediction asserts that for all experimental conditions in which n Achievement or n Affiliation is significantly related to performance, the motive index derived from the first three pictures of the present six-picture form should yield larger correlations than that

TABLE 3

CORRELATION OF ACHIEVEMENT AND AFFILIATION MOTIVE SCORES WITH PERFORMANCE AS A FUNCTION OF PICTURE CUES, PICTURE ORDER, MOTIVATIONAL INSTRUCTIONS, AND INDIVIDUAL VS. GROUP ADMINISTRATION OF PERFORMANCE MEASURES

Motive	Motive Score Source and Performance Measure	Achievement-Orientation				Neutral				Rel.-Affil.	
		Aff. Pictures Shown First (Sections A, F)		Ach. Pictures Shown First (Sections C, S, T)		Aff. Pictures Shown First (Sections A, F)		Ach. Pictures Shown First (Sections C, S, T)		Aff. Pic. First	Ach. Pic. First
		Group (N = 21)	Indiv. (N = 23)	Group (N = 25)	Indiv. (N = 30)	Group (N = 24)	Indiv. (N = 20)	Group (N = 22)	Indiv. (N = 30)	Group (N = 31)	Group (N = 32)
Achieve.	Three aff. pic.										
	Arith.	−.18	.26	−.10	−.16	−.18	−.05	.25	−.08	.02	.44^{b}
	Spatial	−.02	.36^{a}	.04	−.23	−.18	−.12	.16	−.10	−.22	.00
	Matching	−.44	.25	−.20	−.21	−.04	.30	.06	.07	.01	.35^{b}
	Three ach. pic.										
	Arith.	.03	.30	.16	−.12	.08	.38^{a}	.02	.08	.20	.12
	Spatial	.20	.12	.04	.01	−.08	.14	.08	−.12	−.19	−.17
	Matching	.04	.11	.27	−.12	−.12	.10	.27	.01	.02	−.02
	All six pic.										
	Arith.	−.09	.37^{a}	.06	−.18	−.03	.24	.13	.01	.20	.30^{a}
	Spatial	.17	.32	.05	−.11	−.15	.04	.13	−.16	−.25	−.13
	Matching	−.24	.24	.09	−.21	−.11	.23	.24	.05	.02	.15
Affil.	Three aff. pic.										
	Arith.	.25	−.27	−.07	.20	.35^{a}	−.14	.32	.10	.26	.27
	Spatial	.23	−.14	−.02	.11	−.25	−.23	−.07	.34^{a}	.10	−.08
	Matching	.37^{a}	−.16	.12	−.07	.14	−.64	.01	−.05	.00	−.12
	Three ach. pic.										
	Arith.	.62^{d}	.05	.08	−.15	−.19	−.13	−.05	.01	−.25	.14
	Spatial	.00	−.08	−.29	−.32	−.23	.10	.28	.13	−.18	−.13
	Matching	−.13	.02	−.13	−.01	−.41	−.26	.18	.04	−.12	.14
	All six pic.										
	Arith.	.53^{c}	−.22	−.05	.11	.28	−.18	.28	.09	.08	.30^{a}
	Spatial	.17	−.15	−.08	−.04	−.29	−.15	.09	.35^{a}	−.01	−.11
	Matching	.22	−.14	.09	−.07	.00	−.66	.11	−.03	−.06	−.09
	One-tailed Significance Points for Correlation Coefficients										
	^{a}P = .05	.37	.35	.34	.31	.34	.38	.36	.31	.30	.30
	^{b}P = .025	.43	.41	.40	.36	.41	.44	.42	.36	.36	.35
	^{c}P = .01	.50	.48	.46	.42	.46	.52	.49	.42	.42	.41
	^{d}P = .005	.55	.53	.51	.46	.52	.56	.54	.46	.46	.45

based on the last three pictures, regardless of the content of the pictures. Since neither motive index is systematically related to performance under any condition, these data do not permit an adequate test of the prediction. Consider, however, the set of all eight significant correlations between correct scores and motive indices derived from either subset of pictures in either position. The significant correlation is in each case greater than the corresponding one based on the complementary subset of pictures. Three of these eight pairs accord with the prediction, i. e., the larger correlation is based on the subset of pictures presented first; the other five cases contradict it. Finally, as no systematic correlations, either positive or negative, occurred between performance and motive indices, the first prediction also must be considered unsupported.

Further investigation suggests three factors which together may account for the nonreplication in this study of the relationship between n Achievement and performance reported previously (Atkinson & Reitman, 1956).

1. While it is evident from Tables 1 and 2 that for all *S*s combined the level of performance is higher following achievement-orienting instructions than neutral ones, when the com-

parison is made separately by lecture sections, *S*s from Section C show little difference in performance level between the two instructional conditions. This finding may well be a chance phenomenon, but it also may reflect a real difference in motivational level between these *S*s and those from other sections.

2. Were TAT administrations comparably "neutral" in the five different lecture sections? Analysis of variance shows that *S*s from Section S had significantly lower n Achievement scores than *S*s from the remaining four sections.

3. High n Achievement *S*s report significantly more effort than low *S*s to recall the short story during the performance measures under the achievement-oriented group condition. The relatively lesser effort high *S*s consequently expended on the performance measures may account for the absence of the expected motive-performance relations for *S*s who worked under this condition.

It is reasonable to speculate that any or all of these three factors may have confounded a relationship between n Achievement and performance in the groups involved, and to emphasize therefore the one remaining set of correlations relevant to the second prediction which are not subject to these three reservations. For *S*s from Sections A and F who worked under the achievement-oriented, individual condition, the correlation between the six-picture n Achievement index and the arithmetic measure is positive and significant, and the relation of the index to the other two measures is almost as high.

This argument, however, is for two reasons at best only suggestive. It necessarily ignores the over-all chance character of the correlations presented in Table 3, and it depends upon critical variations assumed to exist among *S*s and/or among measurement conditions. While such assumptions are not infrequent in the literature on n Achievement, they consistently frustrate attempts at rigorous integration of the sizable body of research results now available, and thus constitute a compelling argument for far more interexperimental comparability and standardization of method than now exists in this area.

SUMMARY

Three predictions about correlations between n Achievement or n Affiliation and performance under the conditions of this study are tested. (*a*) Significant correlations between motive indices and performance should be positive under all conditions. (*b*) The n Achievement index should correlate maximally with performance under Achievement-Orientation conditions. (*c*) The n Affiliation index should correlate maximally with performance under the Relaxed-Affiliation condition. A fourth prediction, (*d*), stemming from a related earlier study, specifies that significant motive-performance correlations should be largest when the motive index is derived from the first three pictures of the present six-picture measure. The study also evaluates the effects of several other potential determinants of motive-performance relationships.

One of the two forms of a thematic apperceptive measure was administered to each of five groups of *S*s under neutral classroom conditions. All *S*s subsequently participated in a second experimental session 11 weeks later, where they worked at three performance tasks under one of five performance conditions. The Achievement-Orientation conditions yielded a significant performance increment over the Neutral conditions, and this relationship was independent of individual vs. group administration of the response measure, and of differences among *S*s in quantitative and linguistic aptitude. None of the four predictions were supported by the motive-performance correlations obtained. While a supplementary analysis suggests some slight evidence in support of the second prediction, the results and discussion on the whole serve primarily to demonstrate the urgent need for standardized and more precisely controlled techniques in this area.

REFERENCES

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal-attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

BIRNEY, R. C. Experimenter effect on the achievement motive. Unpublished paper, Amherst College, 1956.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

REITMAN, W. R. Motivational induction and the behavioral correlates of the achievement and affiliation motives. Unpublished doctoral thesis, Univer. of Michigan, 1957.

REITMAN, W. R., & ATKINSON, J. W. Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

SPERBER, Z. A study of the role of anxiety level and defense preference in performance under stress. Unpublished doctoral thesis, Univer. of Michigan, 1955.

THURSTON, L. L., & THURSTONE, T. G. *The Chicago tests of mental abilities.* Chicago: Science Res. Associates, 1943.

(Received September 26, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SOME FACTORS AFFECTING THE REDUCTION OF OVERT HOSTILITY[1]

LEONARD BERKOWITZ
University of Wisconsin

As elaborated by Dollard, Doob, Miller, Mowrer, and Sears (1939), the catharsis hypothesis maintains that the occurrence of an act of aggression by an individual reduces the strength of the aggressive drive within him. In accord with this hypothesis, it has been noted that an initially high level of overt aggression toward frustrating agents frequently is followed by a steady decline in this open hostility (Pepitone & Reichling, 1955; Thibaut, 1950), as if the expression of aggression had "drained" the individual's "reservoir of hostility." However, the catharsis hypothesis is not accepted by all psychologists (e.g., Allport, 1954), and the above mentioned negative correlation between initial and subsequent overt aggression is open to a number of different interpretations. Most obviously, the first acts of hostility could have produced guilt or aggression-anxiety which later caused the person to suppress further outward manifestations of this socially disapproved drive.

Whatever the theoretical interpretation, the decrease in overt hostility following an act of aggression has great practical significance. If such a phenomenon can be produced reliably, for example, legal authorities might be advised to provoke expressions of overt hostility in potentially dangerous mobs in situations in which this hostility can be controlled safely in order to prevent later outbursts when the controlling conditions are inadequate.[2] However, such a technique obviously also has great dangers; the first aggressive acts may provoke a later unpredictable and uncontrollable level of violence (Brinton, 1952).

Some conjectures can be advanced as to the conditions leading to each of these reactions. For example, the first aggressive responses are likely to spark later violent outbursts when (*a*) there is a strong aggressive drive (*b*) that had been inhibited by strong external restraints, and others' reactions to the first expressions of hostility reveal not only (*c*) that the restraints are now considerably weakened, but (*d*) that the aggressive behavior is approved by an appropriate reference group. The present study is more directly concerned with factors affecting the reduction of overt hostility. In this case, it can be proposed that the negative correlation between the intensity of early and later overt aggression is likely to arise when (*a*) the individual reacts to his first aggressive acts with guilt or aggression-anxiety,[3] and/or (*b*) the individual is not further angered either by himself or by others. Alternative (*a*) maintains that the guilt or anxiety produced by the initial display of anger can cause the person to inhibit further hostile acts. According to some suggestive findings (Berkowitz, 1958a), middle class college students frequently respond to their own aggressive behavior with aggression-anxiety even when the overt aggression is socially sanctioned. Thus, this proposition does not necessarily assume the validity of the catharsis hypothesis. The first exhibition of aggression may lead to a subsequent decrease in overt hostility, if the aggressive drive is not further heightened or maintained by additional stimulation, because aggression-anxiety inhibits the later manifestations of hostility.

The present experiment is devoted primarily to a test of these conceptions. More specifically, it deals with the effects of events following the individual's initial display of aggression upon his subsequent level of overt hostility. Obviously, further actions by the perceived instigator (*P*) can either weaken or strengthen the individual's hostility producing frustrations, resulting in either less intense or heightened overt aggression. Moreover, the effects of these later actions by *P* can be augmented by self-stimulations. For example, the individual (*S*) may not "get over" his

---

[1] This research was supported by research grant M1540 from the National Institute of Mental Health, United States Public Health Service. Thanks are due to Isabel Birnbaum for her invaluable assistance in collecting the present data.

[2] Ethical considerations aside.

[3] This proposition is not necessarily inconsistent with that advanced by Sears and his colleagues (1953) who suggest that frequent punishment for hostile acts leads to aggression-anxiety which, presumably because this anxiety is frustrating, leads to a heightened aggressive *drive*. The present formulation refers to *overt behavior*.

anger quickly, and could even develop a stronger aggressive drive, if he reminds himself of all the frustrations he has suffered at *P*'s hand. Recalling these frustrations, he stimulates himself to maintain or increase his anger toward *P*. If *S* receives some rewards from *P* after expressing aggression towards him, however, this process of self-stimulation could well increase *S*'s aggression-anxiety. Thinking of *P*'s now "decent" behavior, *S* might feel sorry that he had directed hostility towards him.

Two hypotheses follow from this reasoning: (A) Given that *S* had been angered by *P*, the magnitude of his initial increase in hostility has no relationship to any changes in *S*'s hostility level by the conclusion of the episode (or experiment) if *S* is further frustrated by *P* and is reminded of these frustrations. (B) In contrast, the greater the increase in *S*'s overt hostility toward *P* after the initial instigation the friendlier he should become if the first act of aggression against *P* is followed by *P* rewarding *S* in some manner and relatively frequent recall of this reward.

In addition, the present study seeks to test Feshbach's "symbolic catharsis" hypothesis (Feshbach, 1955). According to this view, symbolic aggression expressed in fantasy activities, such as taking the TAT, results in a cathartic reduction of the aggressive drive. But, as has been pointed out elsewhere (Berkowitz, 1958), there are difficulties confronting this hypothesis. An alternative possibility is that the hostility shown by Feshbach's *S*s in response to the TAT cards gave rise to guilt or aggression-anxiety which produced suppression or repression of further manifestations of this drive. There is little doubt that the arousal of a socially disapproved drive can lead to a decrease in the number of overt drive related responses even in fantasy tasks (Chasdi & Lawrence, 1955; Clark, 1955). Thus, there are two additional questions in the present investigation. Can fantasy activities mediate a reduction in overt hostility, and, if so, is this reduction consistent with the hypothesis that guilt inhibits hostility?

METHOD

Procedure and Experimental Manipulations

Two *S*s, both male or female, were brought together ostensibly for a study of "first impressions." The two *S*s were not given an opportunity to talk to each other at this time. After receiving the instructions and a brief description of their task, they went into separate cubicles where each indicated his impressions of his partner on the first questionnaire (Q1). Each was informed that he would have the opportunity to make other evaluations of him after receiving more information about him later. As a working assumption, it is proposed that expression of attitudes of unfriendliness and rejection on the present questionnaires can be regarded as a form of overt hostility.

When the first questionnaire was completed, the *S*s were told to write two notes to each other, at five-minute intervals, describing themselves. They were to utilize their partner's (*P*'s) self descriptions in evaluating him on the second questionnaire. For the first experimental manipulation in the present 2 × 2 × 2 factorial design, the arousal of Initial Unfriendliness, *E* substituted previously constructed notes for the notes actually written. The notes received by the *S*s in the Initial Unfriendliness condition indicated that *P* had little regard for the University and most of its students, while the *S*s in the Initial Friendliness condition were led to believe that *P* liked life at the University, and was fairly active in student affairs. Following this, *S*s again indicated their impressions of *P* on a second form of the questionnaire (Q2).

The remaining manipulations were undertaken to investigate the effects of further stimulation to aggression after the initial arousal. After *E* had collected Q2 the *S*s were instructed that they would now be given still more information about their partners: they would be told what their partners thought of them. *E* gave each *S* a previously completed Q2, saying that this was the partner's second rating of the *S*. In half of the conditions (Unfavorable Evaluation by *P*) the ratings were uniformly low, averaging 3 on the 10-point scale, while the ratings were uniformly high, averaging 8 on this scale, in the remaining conditions (Favorable Evaluation by *P*). Finally, a set of conditions was established in which the *S*s were to stimulate themselves to further aggression toward *P*. After seeing their coworkers' evaluations of them, half of the *S*s were required to fill out a 145-item adjective check list on which they were to indicate the items that characterized themselves. It was assumed that in completing this task these *S*s would continually remind themselves that they did not like *P*, if they had been "aroused" by him, because *P* did not care for *S* or the University.[4] On the other hand, the *S*s who had received favorable evaluations from *P* supposedly would stimulate themselves to an increased liking for *P*. The remaining half of the *S*s were given a non-self-oriented task taking approximately the same amount of time; they were to respond to four TAT cards. These *S*s, it was assumed, would not stimulate themselves, either to like or dislike *P*, to the same extent as the *S*s given the self-description task. (On the other hand, if Feshbach's (1955) hypothesis of a "symbolic catharsis" is correct, the TAT may show less unfriendliness after anger arousal than the *S*s completing the check list.) Following this task, the *S*s again rated *P* (Q3).

---

[4] The *S*s described themselves instead of *P* for two reasons: (*a*) to remind themselves that it was their self-concept that had been attacked by *P*, and (*b*) unfavorable descriptions of *P* might produce a "cathartic discharge" of hostility.

Dependent Variables and Subjects

All three questionnaires completed by the *S*s contained the same 13 10-point rating scales, but the order of the items differed on Q1. The responses to six intercorrelated items were summed together to yield the unfriendliness score. One of the items reads, "Would you admit him (her) into your circle of close friends?", with the answers ranging from "Definitely yes" (score = 1) to "Definitely no" (score = 10). To summarize, the various ratings of *P* were made at the following times:

After "first impression" (no communication), Q1
After exchanging notes with *P*, Q2
After seeing *P*'s evaluation of them and then responding either to the adjective checklist or the TAT cards, Q3

A second hostility measure was obtained for those *S*s who had been given the four TAT cards. This was a rating on a seven-point scale of the aggression expressed to each card, summed over all cards, using the category system employed by Hokanson and Gordon (1958). The rater did not know the conditions in which the *S*s had been placed. As is the case with the questionnaire scores, the higher the score the greater the hostility.

The *S*s were volunteers from introductory psychology classes at the University of Wisconsin. There are seven women and three men in each condition. The nature of the deceptions and the reasons for them was explained to all *S*s at the completion of the last questionnaire. Most of the *S*s expressed a great deal of interest in the experiment and all readily promised not to talk about the study to other students.

RESULTS

Changes in Unfriendliness

There were no significant differences among conditions, by analysis of variance, in the unfriendliness scores on QI. The means ranged from 19.4 to 24.2. Table 1 presents the mean change in unfriendliness in each of the conditions after the initial arousal (change from QI to Q2), and again after the further stimulation treatments (change from Q2 to Q3). Analyses of variance of these change scores indicate that the initial arousal manipulation was highly successful—there was a considerable increase in unfriendliness from Q1 to Q2 in the Initial Unfriendliness conditions, while the remaining *S*s became somewhat friendlier ($F = 46.46$; $p < .001$). (This was the only significant effect in the variance analysis of the Q1 to Q2 change scores). The *S*s receiving unfavorable evaluations of themselves from *P* also became significantly more hostile from Q2 to Q3 than the *S*s receiving the favorable evaluations. The main effect for Evaluations from *P* is highly significant ($F = 50.89$; $p < .001$), demonstrating the success of the stimulation-from-another treatment.

There also are indications providing at least partial support for the self-stimulation manipulation. First, for the two treatments receiving favorable evaluations from *P*, those *S*s completing the check list, and therefore, who presumably reminded themselves of *P*'s rewards, exhibited a somewhat greater increase in friendliness from Q2 to Q3 than the TAT *S*s. These differences might have become significant if a stronger level of hostility had been aroused in the initial communication period. As will be shown later, the stronger this anger arousal in the check list *S*s (but not in the TAT group), the reliably greater their subsequent change toward friendliness. It may be that the self-stimulation task in this condition caused the *S*s to feel guilty about their previous rise in hostility toward *P*.

In contrast, when *P* had attacked *S* both indirectly through the notes and directly through the unfavorable evaluation (i.e., in the Initial Unfriendliness-Unfavorable Evaluation from *P* condition), completing the adjective check list served to heighten *S*'s unfriendliness toward *P*. Thus, the *S*s describing themselves on the check list became significantly unfriendlier from Q2 to Q3 ($p = .05$ by Duncan multiple range test) than those asked to respond to the TAT cards. As can be seen in the first table, and contrary to the "symbolic catharsis" hypothesis, there was virtually no change from Q2 in this latter group.

From another point of view, Table 1 also suggests that *P*'s evaluation of *S* had its greatest effect on *S*'s attitude towards his "partner" when the evaluation was inconsistent with *P*'s earlier notes. Over all *S*s, the Q2-Q3 attitude changes were greatest when *P*'s evaluations were unfriendly but had been preceded by the friendly notes, or were friendly following the more hostile communications.

This effect is seen more clearly in Table 2. For this analysis, the Initial Unfriendliness Favorable Evaluation and Initial Friendliness Unfavorable Evaluation conditions were combined into a group of *S*s receiving evaluations inconsistent with their prior interactions with *P*. The remaining *S*s comprise a group receiving evaluations that generally could be expected from *P*'s notes. Entered into the table is the number of *S*s in each group whose Q2-Q3 attitude change was consistent with *P*'s evaluation of *S* (i.e., greater friendliness when the evaluation was favorable, greater unfriendli-

TABLE 1

MEAN CHANGE IN UNFRIENDLINESS

P Evaluates *S*:	Initial Unfriendliness		Initial Friendliness	
	Unfavorably	Favorably	Unfavorably	Favorably
Change from Q1 to Q2:				
Adj. C-L *S*s	9.2[a]	9.5	−4.3	−1.2
TAT *S*s	13.1	9.0	−3.2	−3.3
Change from Q2 to Q3:				
Adj. C-L *S*s	12.3c[b]	−10.6b	11.7c	−3.9ab
TAT *S*s	1.0a	−8.2ab	15.4c	0.4a

[a] Positive change scores indicate an increase in unfriendliness while negative scores signify a decrease in this attitude.

[b] Cells containing the same subscript are *not* significantly different from each other, by Duncan multiple range test (Duncan, 1955), at the .05 level. As mentioned in the text, the only reliable difference in the Q1 to Q2 change scores is the main effect for the initial anger arousal.

TABLE 2

NUMBER OF *S*s CHANGING THEIR ATTITUDE TOWARD *P* FOLLOWING *P*'S EVALUATIONS OF THEM, AS A FUNCTION OF THE CONSISTENCY OF THESE EVALUATIONS WITH *P*'S EARLIER COMMUNICATIONS

	Evaluations Consistent with Preceding Notes	Evaluations Inconsistent with Preceding Notes
Number of *S*s exhibiting a change:		
Consistent with *P*'s evaluations of *S*	27	38
Inconsistent with *P*'s evaluations of *S*, or no change	13	2

Note.—$\chi^2 = 9.92$; $p < .01$.

ness when it was unfavorable). As is readily apparent, a significantly greater proportion of the *S*s receiving the inconsistent treatment changed their attitudes in accord with the evaluations.[5] In other words, *S*s were less strongly affected by *P*'s most recent action toward them when this action was consistent with *P*'s earlier behavior. Conceivably, then, the *S*s "shrugged off" the hostile evaluations when these acts followed the unfriendly notes because unfriendly behavior was more or less expected of him.

However, Table 1 also indicates that the presumed self-stimulation process served to increase unfriendliness toward *P* even when *S* had received both unfriendly notes and an unfriendly evaluation from his partner. As noted earlier, the adjective check list *S*s in the Initial Unfriendliness-Unfavorable Evaluation condition exhibited a significantly greater increase in unfriendliness toward *P* from Q2 to Q3 than the TAT *S*s in this condition. Furthermore, the attitude change difference between the favorably and unfavorably evaluated *S*s receiving the hostile notes is statistically significant for the check list *S*s, but not for the TAT group. Thus, it appears that when *S*s are given a non-self-analyzing task (the TAT) after two unfriendly acts from another, they are not particularly affected by the most recent act. They may have adapted to his hostile behavior. The self-stimulation induced by working on a self-description task, on the other hand, does seem to overcome this adaptation effect. What may have happened is that the *S*s, in performing this latter task, "stirred themselves up" to further anger at *P*.

Further information as to the effects of the self-stimulation manipulation is given by the analysis of the self-descriptions themselves. Using only those adjectives whose classification was agreed upon by 7 out of 10 undergraduate judges, it was found that the check list *S*s receiving favorable evaluations from *P* described themselves in a significantly more favorable manner than the check list *S*s given the unfavorable evaluations ($p = .01$ by Mann-Whitney *U* test). Therefore, since the *S*s receiving favorable evaluations were less likely to have their self-enhancement needs frustrated by *P*, it also is likely that they would be more kindly disposed towards him. According to the present assumptions, in completing the self-description task the *S*s then reminded themselves frequently that *P* had either insulted or rewarded them.

Correlations Between Change Scores

The original formulation governing the present study proposed that the increase in unfriendliness from Q1 to Q2 would be negatively correlated with the change from Q2 to Q3, unless *S* stimulates himself to persist or increase in his hostility toward *P*, because the guilt produced by the rise in hostility inhibits further manifestations of aggression. However, *S* may not feel too guilty if the communications received from *P* had not evoked much anger in him. In other words, the strength of the subsequent guilt reaction is a positive function of the degree of anger arousal. It should be possible, therefore, to relate the magnitude of the Q1-Q2 increase in unfriendliness to the Q2-Q3 changes in this attitude within each of the eight experimental conditions. Condition differences in the strength and direction of this relationship can then be traced to situational effects either heightening or weakening the guilt reaction.

---

[5] In this analysis there are no significant differences attributable to whether the change is increased friendliness or increased unfriendliness.

The aggression-inhibition should be weaker, for example, if *S* reminds himself, before completing Q3, of *P*'s attacks upon him. In this case, the guilt feelings either do not arise or they are overcome by the persistent instigation to aggression. Therefore, according to Prediction A, there should be no significant negative correlation for the check-list *S*s in the Initial Unfriendliness-Unfavorable Evaluation from *P* condition. Table 3 reveals that this expectation is fulfilled.

The second specific prediction (B) maintains that the self-stimulation process facilitates the reduction of overt hostility in the *S* receiving a favorable evaluation of himself from *P*. Thus, the check list *S* supposedly would remind himself of his unkind behavior in having "downgraded" *P* when *P* had thought so well of him. Although the self-descriptions give no indication of guilt feelings, there is a significantly negative correlation (tau) between the two sets of change scores for the aroused-then-favorably-evaluated *S*s taking the check list. The greater the increase in hostility after the initial arousal the friendlier the *S*s became after seeing that they had been favorably evaluated by *P*, but only when they presumably reminded themselves of *P*'s high regard for them. The *S*s given the TAT cards did not have this relationship.

TABLE 3

CORRELATION (TAU) BETWEEN CHANGE IN UNFRIENDLINESS FROM QUESTIONNAIRE 1 TO 2 AND CHANGE FROM QUESTIONNAIRES 2 TO 3 ($N = 10$ IN EACH CONDITION)

P Evaluates *S*:	Initial Unfriendliness		Initial Friendliness	
	Unfavorably	Favorably	Unfavorably	Favorably
Q3 after:				
Adj. C-L	+.09	−.64[c]	+.09	−.41[a]
TAT	−.52[b]	−.09	−.24	+.19

[a] $p < .10$.
[b] $p < .05$.
[c] $p < .01$.

TABLE 4

CORRELATION (TAU) BETWEEN CHANGES IN UNFRIENDLINESS AND RATINGS OF TAT HOSTILITY

P Evaluates *S*:	Initial Unfriendliness		Initial Friendliness	
	Unfavorably	Favorably	Unfavorably	Favorably
Change from:				
Q1 to Q2	+.49*	+.02	−.56**	+.53**
Q2 to Q3	−.54**	−.05	+.29	+.54**

* $p < .10$.
** $p < .05$.

Table 3 also presents results which appear to be consistent with Feshbach's "symbolic catharsis" hypothesis. There is a significantly negative rank order correlation between the two sets of change scores for the aroused-and-then-unfavorably-evaluated *S*s who were given the TAT cards. Those *S*s who had the greatest increase in hostility after the initial arousal tended to decrease in unfriendliness after the second arousal treatment, apparently, according to the catharsis hypothesis, as a result of engaging in the fantasy activities.

The correlations between the ratings of TAT hostility and two sets of change scores reported in Table 4 provide further information concerning the hostility reduction process. In particular, they indicate approximately when the *S*s expressing an initially strong dislike for *P* changed in their overt feelings toward him. If the correlation between the Q1 to Q2 change scores and the subsequent fantasy aggression is negative, the reduction in overt hostility obviously occurred between these two sets of responses. On the other hand, if this correlation is nonsignificant and if the correlation between the TAT ratings and the Q2 to Q3 changes is negative, the changes in hostility probably came about some time after the fantasy responses.

As is shown in the table, the reduction in overt hostility after arousal did not occur at the same time in all conditions. In the Initial Unfriendliness-Unfavorable Evaluation condition, for example, there is a tendency for the increase in overt hostility from Q1 to Q2 to be *positively* correlated with the ratings of fantasy aggression. Those who had increased the most in unfriendliness from the first to the second questionnaires were still relatively high in this attitude by the time they responded to the TAT cards. However, there apparently was a reaction to this unfriendliness between the TAT and Q3 so that those *S*s having the highest fantasy aggression tended to become friendlier toward *P*, even though he had evaluated them unfavorably. The five high fantasy aggression *S*s in this condition exhibited, on the average, a 3-unit decrease in unfriendliness, while the five low TAT aggression *S*s increased in hostility toward *P*, on the average, 2.8 scalar units.

These results are to be expected on the basis of the "symbolic catharsis" hypothesis. However, this hypothesis alone cannot predict the results obtained in the Initial Friendliness-Unfavorable Evaluation condition. Here the reaction to the initial expression of overt hostility apparently came prior to the completion of the fantasy task. As is shown in Table 4, those *S*s increasing the most in unfriendliness from Q1 to Q2 tended to have the *lowest* TAT aggression. Clearly, then, the advocate of the catharsis hypothesis must make use of two explanatory mechanisms in accounting for these results: catharsis and some mysterious process that regulates the time of its occurrence. The presumed reaction against one's own unfriendliness did not occur in the "non-aroused" *S*s given the TAT, perhaps because the absence of explicit aggressive responses did not set off the guilt reaction.

DISCUSSION

While the present data do not permit an unequivocal interpretation, it is possible to offer a single explanation for all of the above findings; the view advocated here proposes that each of the negative relationships between initial and subsequent unfriendliness resulted from guilt feelings created by the first hostility reaction.

As indicated earlier, this account in terms of guilt reaction serves as an alternative to Feshbach's (1955) "symbolic catharsis" hypothesis. Where the latter view would maintain that the TAT *S*s in the Initial Unfriendliness-Unfavorable Evaluation condition had reduced their hostility in a fantasy catharsis, the opposing contention presented here argues that the aggressive behavior was merely inhibited because of guilt feelings about aggression. Moreover, this interpretation has the advantage of parsimony. Only a single mechanism, guilt is necessary to account for the present results. In contrast, as was indicated earlier, the adherent to the catharsis hypothesis must utilize two processes in explaining the findings, catharsis and some hitherto unsuspected mechanism governing the time at which this "drainage" will occur.

The differences in the time of the lessening of overt hostility shown in Table 4 certainly appear to be consistent with the guilt arousal explanation. The *S*s displaying hostility reduction relatively early, prior to the TAT, had received only one insult, the unfavorable evaluation, and this evaluation was introduced after *S* was set to be friendly toward *P* by the nonhostile notes. Thus, in the context of the friendly notes, *S* presumably felt guilty for his initial aggression fairly early, perhaps as a direct consequence of his hostile response upon receiving the unfavorable evaluations from *P*. The anger aroused by the evaluation could have been followed by the thought that the evaluation might be justified, ("After all, *P* seems to be a nice guy (or girl)," with the magnitude of the aggression-anxiety initiated by this thought apparently being directly proportional to the increase in hostility toward *P* from Q1 to Q2. In contrast, the *S*s who displayed the hostility reduction following the TAT had been aroused by both the notes and unfavorable evaluations. The hostility arousing notes might have provided a context justifying the aggression felt by *S* after receiving *P*'s evaluation of him, thus delaying the guilt reaction until relatively late.

According to the present interpretation, then, Feshbach's results may be attributable to guilt-arousal in the *S*s given the fantasy task. The aggressive responses made to the TAT cards could have initiated guilt feelings (or aggression-anxiety) within the *S*s for the hostility they felt towards the insulting lecturer, perhaps because in responding to the TAT they became very conscious of this hostility. The *S*s given the nonfantasy task did not have these explicit aggression-cues and therefore did not develop the guilt reaction.

If this reasoning is accepted, it apparently becomes unnecessary to propose that the differences between the adjective check list and TAT *S*s in the Initial Unfriendliness-Unfavorable Evaluation condition stem even partly from the former *S*s' presumed self-stimulation to aggression. As has been just suggested, they may have had a stronger level of overt unfriendliness toward *P* than the TAT *S*s in the same condition, not because they stimulated themselves to aggression, but because the latter *S*s suppressed their overt hostility on Q3.

Nevertheless, there is some presumptive evidence of a self-stimulation process in the *S*s receiving the favorable evaluations from *P*. In the case of the *S*s taking the check list in these conditions, the greater their increase in unfriendliness from Q1 to Q2 the friendlier they became from Q2 to Q3. It may be that the relatively hostile *S*s in these conditions, in responding to the self-description task, reminded themselves of their "unkindness" in having been unfriendly toward *P* when he had thought so well of them. They, therefore, may have stimulated themselves to increased guilt feelings which produced the manifestations of friendliness on Q3. Thus, if this self-stimulation could have occurred in the *S*s receiving the favorable evaluations, it also may have occurred in the *S*s getting the unfavorable evaluations.

In conclusion, a final point may be made concerning the nature of the frustrations giving rise to aggression. A number of writers have suggested (Berkowitz, 1958, p. 270) that an individual is most likely to become angered by a frustration when this frustration involves the nonfulfillment of an expectation. Data reported earlier provide some support for this contention. Assuming that both the unfriendly notes and the unfavorable evaluations from *P* were frustrating to *S*, the second type of frustration followed the earlier frustration for some *S*s but not others. However, instead of the effects of the two frustrations summating, as Dollard, Doob, Miller, Mowrer and Sears (1939) have predicted, the greatest unfriendliness provoked by the unfavorable evaluation over all *S*s came about when this frustration was inconsistent with *P*'s earlier behavior, (i.e., had been preceded by the relatively nonfrustrating friendly notes). Conceivably, the *S*s receiving this inconsistent treatment had expected friendlier behavior from *P*, so that the nonfulfillment of this expectancy gave rise to a relatively strong aggressive reaction. Moreover, this may be a special case of a more general principle. *P*'s evaluations generally produced the greatest over all change in *S*'s attitude towards him, whether this was an increase in friendliness or unfriendliness, when the nature of the evaluations presumably was unexpected.

SUMMARY

Members of pairs of *S*s, ostensibly engaging in a study of "first impressions," indicated their degree of liking for each other at various times during a 50-minute experimental session, right after first seeing each other (Q1) and then twice more after instigations to aggression. Three sets of instigating conditions were created in a 2 × 2 × 2 factorial design. First, half of the *S*s were caused to have an initial dislike for their partner (*P*) by manipulating the written communications supposedly sent by him. The remaining *S*s received "friendlier" messages. The *S*s then completed the second questionnaire (Q2). Following this, half of the *S*s were told that *P* had evaluated them unfavorably on this latter questionnaire, while the other *S*s were led to believe that *P* had evaluated them favorably. But before the *S*s were given the final attitude questionnaire (Q3), half were asked to respond to four TAT cards, with the remaining *S*s assigned to describing themselves on a long adjective check list. It was assumed that the latter *S*s would remind themselves of *P*'s behavior toward them and thus would stimulate themselves to an increased liking or disliking for *P*.

The results indicate the following:

1. Comparisons among the condition means indicates that *P*'s second act, his evaluation of *S*, had its greatest effect in changing *S*'s attitude toward *P* when this act was inconsistent with *P*'s earlier behavior. Thus, if the evaluation was friendly (favorable) but had been preceded by unfriendly communications from *P* there was more likely to be an increase in friendliness toward *P* than if this favorable evaluation had been preceded by the hostile notes. Similarly, the unfriendly (or unfavorable) evaluation was more likely to produce an increase in anger when the earlier communications from *P* had been friendly than when these messages had been hostile in tone. Since the *S*s probably expected evaluations from *P* consistent with his earlier behavior, it is suggested that a frustration is most likely to

give rise to aggression when this frustration involves the nonfulfillment of an expectation. There also is evidence that the presumed self-stimulation induced by working on the self-description task was effective in further increasing unfriendliness toward *P*, as a result of his unfavorable evaluation of *S*, even when this had been preceded by the hostile communications from *P*.

2. The *S*s receiving favorable evaluations of themselves from *P* showed a negative correlation between increase in unfriendliness from Q1 to Q2 and changes in this attitude from Q2 to Q3, regardless of whether they had been initially angered or not, if they took the adjective check list but not if they responded to the TAT. It is hypothesized that the *S*s in these conditions displaying the greatest increase in hostility after their first impressions, in performing the self-description task, reminded themselves of their unkindness in having been unfriendly toward *P* when he had thought so well of them (as indicated by his Q2 ratings). This self-stimulation to strong guilt presumably led to manifestations of friendliness on Q3.

3. The *S*s given the intervening TAT task showed a negative relationship between the increase in unfriendliness from Q1 to Q2 and fantasy aggression if they had been angered only once (by the unfavorable evaluations of themselves). It is suggested that these *S*s felt guilty for their unfriendliness relatively soon, prior to taking the TAT, so that their guilt produced an inhibition of hostility on the TAT. In contrast, the *S*s angered by both the notes and unfavorable evaluation from *P* showed a negative relationship between fantasy aggression and the subsequent attitude changes from Q2 to Q3. Here it is conjectured that these *S*s did not feel guilty as soon as the other *S*s because the first hostile notes from *P* to some extent served to justify the anger provoked by the evaluation, so that the guilt reaction inhibiting overt manifestations of unfriendliness did not occur until just before the final questionnaire was completed.

REFERENCES

ALLPORT, G. W. *The Nature of Prejudice.* Cambridge, Mass.: Addison-Wesley, 1954.

BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, **55**, 257–283.

BERKOWITZ, L. Manifest hostility level and hostile behavior. *J. soc. Psychol.*, in press.

BRINTON, C. *The Anatomy of Revolution.* New York: Prentice-Hall, 1952.

CHASDI, ELEANOR H., & LAWRENCE, MARGARET S. Some antecedents of aggression and effects of frustration in doll play. In D. McClelland (Ed.), *Studies In Motivation.* New York: Appleton-Century-Crofts, 1955.

CLARK, R. A. The effect of sexual motivation on phantasy. In D. McClelland (Ed.), *Studies In Motivation.* New York: Appleton-Century-Crofts, 1955.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and Aggression.* New Haven: Yale Univ. Press, 1939.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–45.

FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3–11.

HOKANSON, J. E. & GORDON, J. E. The expression and inhibition of hostility in imaginative and overt behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 327–333.

PEPITONE, A. & REICHLING, G. Group cohesiveness and the expression of hostility. *Hum. Relat.*, 1955, **8**, 327–337.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.

THIBAUT, J. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, **3**, 251–278.

(Received October 3, 1958)

g - 426
8. 9. 70

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

FANTASY RESPONSES TO SUBLIMINAL STIMULI

MICHAEL J. GOLDSTEIN AND RICHARD P. BARTHOL[1]
University of California, Los Angeles

RESEARCH on subliminal stimulation has been largely directed toward showing that social, situational, or personality factors influence a subject's (*S*'s) ability to decode stimuli ordinarily considered as below threshold. The task of *S* typically is to decode briefly presented visual or auditory stimuli in a situation in which there is reason to believe that the stimuli will have differential impact on the *S*s. Since the method of study uses impoverished stimuli that are gradually increased in clarity, conventional psychophysical techniques have been employed. The conclusion that social, situational and personality factors do influence perception has been dependent to a large degree on the appropriateness of these traditional psychophysical methods to the problem at hand. The analysis of Eriksen (1958) suggests that the use of these conventional psychophysical methods makes the motivational interpretation suspect. This argument has been underscored by Goldiamond (1958) who, although writing from a different point of view, presents some very convincing evidence that the response chosen as the indicator of perception has more to do with the obtained results than does the motivational character of the stimuli.

This criticism of methods has stimulated the development of another, less controversial approach to the problem of personality variables in perception. The latest approach, derived largely from the work of Klein, Spence, Holt, and Gourevitch (1958), involves a basic change in the experimental task assigned to *S*. Instead of requiring active decoding of a subliminal stimulus, *S*s are instructed to respond to a clearly supraliminal stimulus. While *S*s respond in this manner, *E* periodically introduces a subliminal or more accurately, incidental stimulus, and notes the effect of the stimulus upon the response to the supraliminal stimulus. If the original methods of study, using threshold measurement, study perceptual learning, then Klein's technique appears to shift the emphasis to the study of *incidental* perceptual learning. The results with this technique suggest that it may redirect the emphasis in perception-personality research to more fruitful ground.

The present paper is concerned with the effect of incidental stimulation upon fantasy. Since it is assumed that incidental stimuli affect behavior by registering outside of awareness, it was hypothesized that such stimuli would most likely influence a response such as fantasy which also is believed to arise from stimuli outside of awareness. Specifically, we predicted that the emotional tone of *S*s' fantasy responses would parallel the emotional tone of the incidental stimulation.

STUDY I

Method

The present report is based upon two consecutive studies which were of identical design except for one major variable described below. *S*s were volunteers from the introductory psychology course at UCLA who were satisfying a course requirement. Thirty-six *S*s were divided randomly into three equal groups which differed in the type of incidental (subliminal) stimulation presented. A *positive* group received four positively toned words, a *negative* group received four negatively toned words, and a *control* group received nonsense symbols which resembled words when projected rapidly. As the supraliminal fantasy stimulus, four slides made from TAT cards were used. Table 1 presents the words received by the three groups and the TAT cards with which they were projected.

The three basic groups were subdivided into a balanced latin square design with two replications, making it possible to evaluate the main experimental effect as well as any interactions between order of presentation and type of TAT card. *S*s were instructed to make up a story about each TAT picture following the instructions developed by McClelland (1953). These instructions provided a 20-sec. delay between

---

[1] An abridged version of this paper was originally presented by the senior author as part of the symposium "Subliminal Stimulation and the Social Scientist" at the 1958 APA convention. The present studies were conducted on a grant from the Faculty Research Committee of the University of California at Los Angeles. The authors would like to express their appreciation to Allen E. Edwards, Barbara Rosenquest, Samuel Himmelfarb, and Sue Smith for their assistance in various stages of the studies.

TABLE 1

SUPRALIMINAL TAT CARDS AND INCIDENTAL WORDS PRESENTED CONCURRENTLY, BY GROUP

Group	N	TAT Card Number (Supraliminal)			
		1	3GF	7BM	10
Positive	12	Success	Laugh	Happy	Love
Negative	12	Failure	Cry	Angry	Hate
Neutral	12	////	////	////	////

the presentation of the TAT stimulus and the initiation of a response by *S*. A "dummy" TAT slide, upon which no incidental stimulus was projected, started the series. This was followed by the four experimental cards with overlaid incidental stimuli. *S*s told their stories which were picked up by a hidden tape recorder. Following the four critical slides, *S*s were informed about the incidental stimulation, and two tests were made of the recall of the incidental stimuli. The TAT cards were reprojected one at a time and *S*s were instructed to guess the stimulus words (*a*) on a free-recall basis and (*b*) from a four-word forced-choice test. In the first study, one *S* spontaneously reported seeing the words during the experimental procedure and was replaced by another *S*. Almost all *S*s expressed amazement when informed of the presence of the incidental stimulation.

Method of incidental stimulation. *S*s tested individually were seated 8 ft. from a screen. Two adjacent projectors presented the TAT slides and the incidental stimulation. The projectors were set 4 ft. to the rear and slightly to the left of *S* behind a burlap curtain that hid them from view. The TAT slides were presented in a normal fashion without interruption. The duration of the incidental stimulation was controlled by a large (4 ft. diameter) episcotister which was connected to a constant-speed electric motor with a variable speed transmission. Varying the speed of the motor and the size of the slit in the episcotister gave a wide variety of exposure times. Since *S*'s task was not to actively decode the incidental stimuli, the usual methods of threshold determination were not appropriate. Therefore, settings for the incidental stimulation were determined in a pilot study with a different group of *S*s who were actively engaged in responding to a TAT slide. It was found that an exposure of 3.2 msc. did not distract *S* from making up his story. Neither spontaneous verbalizations nor detailed questioning of *S*s revealed that they had any awareness of the presence of the words overlaid on the TAT slides, although the material was visible to *E*s. This setting was then used for both studies reported in this paper. It was not found necessary, contrary to expectation, to vary the setting on an individual basis. The same setting was thus used for both studies. The incidental stimuli were presented repetitively at a 1/sec. rate providing 20 exposures while each TAT slide was on the screen.

TABLE 2

ANALYSIS OF VARIANCE OF EMOTIONAL TONE RATINGS FOR CLEAR STIMULI

Source	Sum of Squares	*df*	*F*	*P*
Total	117.50	143		
Between Subjects				
Treatment	1.52	2	1.132	N.S.
Order	1.09	3	0.540	N.S.
O × T	5.12	6	1.271	N.S.
Error	16.12	24		
Within Subjects				
Trials	0.13	3		N.S.
Cards	36.04	3	21.881	<.001
Treat. × Trials	2.63	6	1.321	N.S.
Ord. × Treat. × Trials	12.01	18	1.214	N.S.
Error	42.84	78		

Results

The TAT stories were collected by a research assistant and neither of the authors had any direct contact with *S*s. All stories were typed from the tapes with no identifying marks. Ratings of the emotional tone of the stories were then made independently by the authors using Eron's (1950) general rating scale for the emotional tone of TAT stories. The ratings of the two authors correlated .79. Using the mean of the two ratings on each *S*, an analysis of variance was performed, the results of which are shown on Table 2.

It can be seen from the table that the presence of the incidentally presented words failed to influence the emotional tone of *S*s' fantasy productions. The only significant *F* present in the data was between TAT cards, supporting Eron's (1950) earlier finding that the different TAT cards are far from ambiguous in emotional tone.

Since it is possible that a message could be received and yet have no effect upon behavior, an analysis was made of the free verbalizations and the forced-choice responses of *S*s to determine if any stimulus registration had taken place. Using Fisher's exact probability method, an analysis was made of the frequency of correct and incorrect guesses of the incidental stimuli under free recall and forced-choice conditions. The results, by individual card, and over all TAT cards, failed to indicate superior guessing in the two experimental groups when

TABLE 3
ANALYSIS OF VARIANCE OF EMOTIONAL TONE RATINGS FOR AMBIGUOUS STIMULI

Source	Sum of Squares	*df*	*F*	*P*
Total	97.79	143		
Between Subjects				
Treatment	7.41	2	4.333	<.025
Order	.53	3	0.019	N.S.
O × T	3.71	6	0.723	N.S.
Error	20.53	24		
Within Subjects				
Trials	.77	3	0.003	N.S.
Cards	16.46	3	10.894	<.001
Treat. × Trials	3.46	6	1.145	N.S.
Ord. × Treat. × Trials	5.65	18	0.624	N.S.
Error	39.27	78		

compared to the control group. Even when the free-verbalizations and forced-choice responses were classified merely on the basis of whether they were positive or negative in emotional tone, no differences were found.

STUDY II

The large F between the TAT cards suggests that the stimulus properties of the cards inhibit variability among *S*s. It seemed possible that the fantasy stimuli would have to be much more ambiguous if the incidental stimulation were to have any effect. Therefore, a second study was conducted in which the ambiguity of the TAT cards was markedly increased by adjusting the lens on the TAT projector to maximum diffusion of the projected image. Although it is difficult to describe the resulting pictures quantitatively, qualitatively there was an increased fuzziness of the picture, and a general reduction of clarity of the figures and the nature of their activity. However, it was still clear that the pictures dealt with human-like figures. This technique risked the creation of a more homogenous visual field against which the incidental stimuli might have been more visible. However, the subsequent analysis of the free-verbalization and forced-choice responses did not indicate greater awareness of the incidental stimuli when compared with the clear stimulus group.

Aside from the variation in ambiguity the second study was identical in every way to the first. A second sample of 36 *S*s, drawn from the same population, was used.

Results

The stories in Study II were analyzed in the same way as in the first study. Table 3 below presents a summary of the analysis of variance of the ratings of emotional tone.

It can be seen in Table 3 that the predicted difference between the three treatment groups is statistically significant ($F = 4.333$, 2 and 24 df, $p < .025$). An analysis of the individual means reveals that the positive group gave the most positively toned stories, the negative group gave the most negatively toned stories, and the control group lay between the two. These results occurred in spite of the fact that the increase in stimulus ambiguity did not eliminate the variance due to TAT stimulus characteristics. Although the obtained F of 10.894 between TAT cards is not as large as in the first study, it is still highly significant ($p < .001$ for 3 and 72 df). Again the implication is that structural characteristics of a TAT card place restrictions upon story tone and that incidental stimuli can influence the story tone only within these restrictions and cannot produce any gross modifications of fantasy. For example, when a negatively toned TAT card was presented, the positive group gave the least negatively toned stories, the negative group gave the most negatively toned stories, and the control group fell in between. A negatively toned incidental stimulus could not stimulate a negatively toned story to a positively toned TAT stimulus, merely a less positively toned story than might ordinarily be told.

It was expected that the presence of differences in the treatment groups would be paralleled by an increase in *S*s' awareness of the nature of the incidental stimuli. The data were to the contrary, however. Fisher's exact probability method applied to the frequency of correct and incorrect guesses in free recall or in the forced-choice test failed to indicate superior guessing of the incidental stimuli in the two experimental groups when contrasted with the control group.

DISCUSSION

The results of the two studies indicate that the presence of incidental (subliminal) stimuli does not produce a uniform effect upon fantasy responses. When the fantasy stimulus is highly structured and present sufficient stimulus clues for its organization, incidental stimuli are of little consequence. However, as the number of cues available for organization of the fantasy stimulus are reduced, incidental stimuli become part of the effective stimulus complex. It is as if an organism, in reacting to its environment, responds to the hierarchy of perceptual cues in some sort of selective order, with the most prominent and familiar being high in the cue hierarchy. However, when the perceptual cues that are high in the hierarchy are not sufficient for the organization of the stimulus situation, cues of lesser intensity are employed. A visual perception task may be thought of as a special case of any message decoding problem. In certain instances, the manifest content of a written message contains sufficient information for reliable, although not necessarily valid, decoding. Other written messages are more ambiguous and may require information about the writer, such as his sex, age, beliefs, group membership, etc. for decoding. Incidental stimuli appear to provide that kind of information when the more prominent stimuli are insufficient for reliable decoding.

The failure to find differences in free-recall and forced-choice recognition in the second study raises some difficult questions. It was assumed, in the design of the present study, that discrimination of the incidental stimuli must precede any influence upon fantasy. Yet *S*s were unable to give evidence for this sort of discrimination following the fantasy task. Two possibilities arise from this finding. First, it is possible that incidental stimuli cause a response while present on the screen but that they have little sustained effect. A study in progress is designed to provide an answer to this question by determining the effects of incidental stimulation upon the later learning of the same material presented clearly. If material presented as incidental stimuli register outside of awareness, then the future discrimination of these same stimuli should be facilitated when they are clearly visible.

The second possible explanation for the failure to find evidence of word discrimination is that actual discrimination of word form did not take place. It has been suggested by Osgood (1957) in his analysis of motivational factors in perception that "the 'hint' provided by the incomplete stimulus display in a tachistoscope may be sufficient to excite the dominant evaluative component of the total mediational process without being sufficient to excite the total reaction pattern." If this analysis is correct, *S*s in the present study were reacting emotionally to the incidental stimuli and were not making a discrimination of a more precise nature. Thus, the same effect would be obtained whether the specific word projected in conjunction with a TAT card was *success* or any other word with a similar positively toned quality (love, happy, delight) were projected. An analysis of free-recall and forced-choice responses of *S*s failed to indicate that this type of gross evaluative discrimination was taking place. However, since these responses were obtained on a delayed basis, their value in a test of this hypothesis may be questioned. Further research is needed using a broader sample of incidental stimuli of varying evaluative significance before the second hypothesis can be adequately tested.

The success of the method used in the present study suggests that it would be fruitful for further experimentation on problems of personality and perception. The question is no longer "Does *S* see or not see a particular stimulus?" It becomes: "What are the effects of stimulation of differing degrees of clarity or intensity upon subsequent behavior?" As so many philosophers of science have indicated, a change in the way of asking a scientific question frequently results in increasing the probability of finding a meaningful answer.

SUMMARY

Two studies were conducted to determine the effect of incidental visual stimulation (words) upon fantasy responses. Both studies used four TAT slides as the supraliminal stimuli. Three experimental groups were formed according to the emotional tone of the words used as incidental stimuli: positive group, negative group and control (neutral)

group. The control group received nonsense symbols that resembled words.

The results of the first study failed to indicate that the incidental stimuli had any effect upon the emotional tone of the fantasy responses. The data did reveal that the different TAT cards were clearly not ambiguous stimuli and tended to elicit stories of a characteristic emotional tone.

In the second study, the ambiguity of the TAT slides was markedly increased. Analysis of the stories under this condition showed that the incidental stimuli altered the emotional tone. The positive group produced the most positively toned stories, the negative group produced the most negatively toned stories, and the control group produced stories of an intermediate emotional tone. The emotional impact of the different TAT cards was still stronger than that of the incidental stimuli. The analysis of *S*s' guesses about the nature of the incidental stimuli failed to indicate discrimination of these stimuli.

The results are interpreted in terms of an information decoding model in which perceptual cues of differing intensities are used by *S* in an ambiguous situation. Where the message is unambiguous, *S*s react only to the prominent cues in the stimulus situation; however, when a message is ambiguous, *S*s are likely to respond to more peripheral or incidental aspects of the stimulus complex.

The question of whether *S*s are discriminating the full meaning of a message or merely reacting to the evaluative (emotional) aspect of a message is also discussed.

REFERENCES

ERIKSEN, C. W. Unconscious processes. In M. R. Jones. (Ed.), *Nebraska symposium on motivation, 1958*. Lincoln: Univer. Nebraska Press, 1958.

ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64** (9, Whole No. 315).

GOLDIAMOND, I. Subliminal perception: I. Indicators of perception. *Psychol. Bull.*, 1958, **55**, 373–411.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNAH. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

MCCLELLAND, D. *The achievement motive.* New York: Appleton-Century-Crofts, 1953. P. 98.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1957. Lincoln: Univer. Nebraska Press, 1957.

(Received November 28, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE ACHIEVEMENT MOTIVE, GOAL SETTING, AND PROBABILITY PREFERENCES

JOHN W. ATKINSON,[1] JARVIS R. BASTIAN, ROBERT W. EARL,[2] AND GEORGE H. LITWIN[3]
University of Michigan

THIS investigation deals with the effect of individual differences in strength of achievement motive on risk taking behavior. It focuses upon (*a*) goal setting as inferred from choices individuals make among tasks that differ in difficulty (see Lewin, Dembo, Festinger, & Sears, 1944), and (*b*) the preferences of individuals in an imaginary gambling situation for options that are equated for expected monetary value but differ in the stated probability of winning (see Edwards, 1953).

The senior author (1957, 1958b) has presented a theoretical model that extends some ideas developed earlier by Escalona and Festinger in the resultant valence theory of level of aspiration (see Lewin, et al., 1944). It accounts for the association between achievement motivation and entrepreneurial activity involving risk taking highlighted in the writings of McClelland (1955, 1958). The theoretical model was initially formulated to explain the results of two experiments dealing with different aspects of risk taking behavior. Atkinson (1958a) found that the level of performance of female college students was significantly higher when the probability of winning a small monetary prize by getting a high score was 1/2 than when the expectancy of winning was either very high (3/4) or very low (1/20). McClelland (1958) found that children in kindergarten and second grade who were judged to be highly motivated to achieve on a graphic expression measure of n Achievement preferred an intermediate degree of risk (or difficulty) more frequently in several different competitive games than children presumed to be low in n Achievement. The latter group more often than the high n Achievement group favored playing it very safe or being very speculative.

The principle of motivation evolved out of these and a series of earlier experiments on n Achievement and performance (Atkinson & Reitman, 1956; Atkinson, 1958b) is very similar to principles proposed by Tolman (1955) and Rotter (1954): Motivation = *f* (Motive × Expectancy × Incentive). However, it differs from some similar formulations in its conception of a motive as a relatively stable disposition of the personality.

It is assumed that in competitive achievement situations, two different motives are invariably aroused by cues that elicit expectancies (subjective probabilities) of success and failure—the achievement motive (n Achievement) and the motive to avoid failure. It is further assumed that the incentive value of success, that is, the amount of satisfaction inherent in successful accomplishment of a particular task, is an inverse linear function of the strength of the expectancy of success. Incentive value of success is assumed to equal $1 - P_s$ where P_s is the subjective probability or measure of the strength of the expectancy of success. Hence for a given individual, the multiplicative combination of Motive to achieve success, Expectancy of success (P_s), and Incentive value of success ($1 - P_s$) will always produce a maximum value when P_s is .50, as in tasks of intermediate difficulty.

It is assumed that the (negative) incentive value of failure is greater the easier the task (specifically, that $I_f = -P_s$). That is, the shame and humiliation accompanying failure of an easy task is assumed always to be much greater than that accompanying failure of an admittedly very difficult task. Given this assumption it can be shown that the multiplicative combination of Motive to avoid failure, Expectancy of failure, and Incentive value of failure also reaches a maximum value for a given individual when P_s is .50. However, in this instance the product of the three variables, which represents the intensity of aroused motivation, has a negative sign implying anxiety and avoidant behavior. If both the tendency

---

[1] This research was supported by a Ford Foundation grant to the senior author for studies of Personality Dynamics. Preliminary results were presented in a paper at the 1957 APA meeting in New York.

[2] Now at Claremont Graduate School.

[3] Now at Harvard University.

to approach success and the tendency to avoid failure are simultaneously aroused, and if the approach-avoidance conflict that ensues is resolved algebraically as supposed, the behavior of an individual when confronted with an intermediate degree of risk will clearly depend upon which of these two motives is stronger (see Atkinson, 1957, 1958b).

It is assumed that persons having high n Achievement scores as measured by thematic apperception (McClelland, Atkinson, Clark, & Lowell, 1953; Atkinson, 1958b) or by the French Test of Insight (French, 1958), a modified apperception test, are more likely than persons having low scores to be persons in whom the achievement motive is stronger than the motive to avoid failure. Consequently, it is predicted that persons having high n Achievement scores will prefer intermediate risk (or difficulty) to a greater extent than persons having low n Achievement scores.

The prediction of risk taking behavior of particular persons should obviously be improved when the strength of both the motive to achieve success and the motive to avoid failure are assessed and taken into account. Low n Achievement scores can hardly be viewed as a measure of the strength of the motive to avoid failure. Nevertheless, there is considerable presumptive evidence that the motive to avoid failure is likely to be stronger than the achievement motive in many persons having low n Achievement scores. Atkinson (1953) found that persons having low n Achievement scores show a decreasing tendency to recall interrupted tasks that might be construed as past failures when the instructions for performance of the tasks become more achievement related. This effect has often been called repression of failure. McClelland and Liberman (1949) and Moulton, Raphelson, Kristofferson, and Atkinson (1958) have demonstrated the phenomenon commonly called perceptual defense among *S*'s having low n Achievement scores when presented with failure words tachistiscopically. This result is particularly evident immediately after they have taken some kind of achievement test. Brown (1953) has found that persons in the low n Achievement group are more rigid in solving the *Einstellung* water-bottle problems under achievement test conditions. Raphelson (1957) has found that persons having low n Achievement scores *obtained from thematic apperceptive stories during a stressful achievement test* score higher in Test Anxiety (Mandler & Sarason, 1952) and show a greater amplitude of PGR in the stressful achievement situation then persons having high n Achievement scores. Taking all of this behavioral evidence into account, it is not unreasonable to assume that the *unmeasured* fear of failure is probably relatively stronger in relation to strength of achievement motive in more persons having low than high n Achievement scores.[4] Hence there is some basis for predicting a tendency for a group of persons low in n Achievement to avoid intermediate degrees of risk.

METHOD

Subjects

The data were collected in five laboratory sections of a sophomore-junior level psychology course at the University of Michigan during the fall of 1956. The students in the course were the *S*s and the instructors served as the *E*s. In the present analysis only the data for males will be examined, as the n Achievement measure for females has always constituted a separate problem (see McClelland, et al., 1953 pp. 172–181; Atkinson, 1958b). There were 66 male *S*s, not all of whom, however, were available for all of the separate tests of risk taking behavior.

Measure of n Achievement

Several weeks prior to the experimental session the *S*s were given the Test of Insight developed by Elizabeth French (1958). The test consists of 10 short verbal descriptions of behavior that *S*s are asked to analyze. An example cue is, "Bill always lets the 'other fellow' win." Both Forms I and II, as described by French, were used. Half of the *S*s were given each form. The stories were scored for n Achievement according to the procedure described by McClelland, et al. (1953; also in Atkinson, 1958b) by an experienced judge whose reliability had been established as above .90 on earlier samples.[5] The separate distributions of the male n

---

[4] Three subsequent studies of achievement motivation and risk taking behavior have used an independent measure of what is here conceived as the motive to avoid failure, the Mandler-Sarason Test Anxiety Questionnaire (1952). The results of these studies provide additional support for the inferences made here concerning fear of failure among the *S*s having low n Achievement scores (see Atkinson & Litwin, 1959; Litwin, 1958; Vitz, 1957).

[5] We wish to acknowledge the assistance of Joan Munson, who scored the protocols. Because so many of the interpretations had no achievement imagery, the distinction between doubtful (TI) and unrelated imagery (UI) was ignored, and unrelated imagery was scored 0 instead of -1.

Achievement scores for the two forms of the Test of Insight were broken at the median and those above the median designated as high n Achievement, those below the median as low n Achievement.

Goal Setting in a Modified Shuffleboard Game

In the experimental session each class (*N*s = 15 to 20) was shown a shuffleboard arrangement. Separate arrangements were set up for males and females. A chalk circle one ft. in diameter was marked on the floor. Fifteen lines, one ft. apart, were marked on the floor in chalk. The closest line was 1 ft. away from the target; the farthest line was 15 ft. away. A chalk number was written beside each line. The number corresponded to the distance from the target.

The instructions given to the class were:

> Today we are planning to play a little shuffleboard to see how good you are at it. Each of you will be given five practice shots. You can take your shots from any line you want and you can move around if you care to.

Then each *S* took five shots with other members of the class looking on. The men shot on one side of the room, the women on the other. There was a good deal of informal banter during the test period. No attempt was made to interfere with or to control the lifelike game atmosphere that prevailed. However, one person did keep track of the shots and hits. When *S* finished he wrote his experimental code number on the sheet recording his shots.

Controlled Practice

The *S*s were then sent out of the room and brought back one at a time for controlled practice. Each *S* was given 10 practice shots from Lines 1, 3, 4, 5, 6, 7, 9, 11, 13, and 15 in a predetermined random order. One male and one female *S* and two recorders were present during this regulated practice. As before, male and female *S*s were run on opposite sides of the classroom.

Estimate of Subjective Probability of Success and Final Test

After all *S*s had finished the practice shots they were again brought in individually and told:

> Now comes the big game to see how good you really are. You will have five shots and can stand anywhere you please. Try to get the highest score. Your score will be the sum of the distances of your hits. Before you begin *we would like your estimate of how many of the 125 people in the class you expect to beat.* (These estimates were recorded.) We have marked the chance of hitting at each of four of the lines. These probabilities are based on the practice shots of all class members.

The following probabilities were written on the floor next to the lines as listed: 2 ft.—.60, 4 ft.—.50, 10 ft.—.30, 15 ft.—.02.[6] Since two games, one for males and one for females, were being played simultaneously, the *E* recorded the data for the males and designated one of the *S*s who had finished to record the data for the females.

Betting Preferences

After the shuffleboard game, the *S*s were asked to indicate their preferences among a series of imaginary bets. A questionnaire was passed out containing 15 paired bets and the *S* was asked to select one bet from each pair. The bets differed in probability of winning but were all of a constant expected monetary value. The probabilities of winning that were used were 6/6, 5/6, 4/6, 3/6, 2/6, 1/6. Two different forms of the betting preferences were used, half of the *S*s getting each form. One form had small value bets; the expected value of these bets was 30¢. The other form had large value bets; the expected value of these bets was $300. A sample pair of bets from the small form is: 3/6 to win 60¢ vs. 5/6 to win 36¢. The 15 pairs represented all possible pairs of the six bets. The brief instruction asked *S*s to imagine that they were in a gambling situation with friends shouting encouragement while they rolled a single die. The stated options were interpreted for them in the instructions, e.g., 3/6 meant if numbers 1, 2, or 3 come up on the six-sided die, you win, etc.

Measures of Subjective Probability of Success

No direct measure of expectancy or subjective probability was available for the shuffleboard game. For the final test shots, the probabilities written on the floor and represented as the practice probabilities of the *S* were taken as the measures of expectancy. However, for the free practice shots not even these were available. Geographic distance from the target seemed the most reasonable and obvious measure of expectancy. To predict from the theoretical model, it is necessary to make the following assumptions: (*a*) subjective probability of success at the one ft. line is nearer to 1.00 than .50; (*b*) subjective probability of success at the 15 ft. line is nearer to 0 than .50. These assumptions justify dividing the 15 lines into equal thirds corresponding roughly to high, intermediate, and low subjective probability of success. For the betting preferences, the stated objective probabilities may be considered the best estimate of expectancy of success.

As a result of a misunderstanding on the part of several of the persons recording the data in the shuffleboard game, only partial data were available for some *S*s. The possibility of using only *S*s on whom all data were available was rejected since it would have resulted in very small *N*s for some comparisons. Hence, all available data were used, even though slight variations in the size of samples resulted for the various aspects of the shuffleboard data. For the same reason, the sample for the shuffleboard analysis is somewhat

---

[6] The numbers did not accurately represent the obtained probabilities. On the basis of pretesting it was anticipated that the obtained probabilities would range from above .80 to below .20. These obtained probabilities were to be written on the floor. However, the *S*s did so poorly in the controlled practice that even the closest lines had probabilities of hitting that were less than .50. The probabilities that were written on the floor were, as a result, a compromise between what had actually occurred (and hence what would be believed by the *S*s) and the range of difficulty that was expected when the experiment was designed.

TABLE 1

PERCENTAGE OF *S*s HIGH ($N = 27$) AND LOW ($N = 31$) IN n ACHIEVEMENT TAKING SHOTS AT VARIOUS DISTANCES FROM A TARGET IN A SHUFFLEBOARD GAME

Trial	n Achievement	Distance from Target			Intermediate vs. Very Easy and Very Difficult Combined
		Very Easy 1–5 ft.	Intermediate 6–10 ft.	Very Difficult 11–15 ft.	
1	High	4	56	41	$\chi^2 = 8.18^{**}$
	Low	26	19	55	
2	High	4	56	41	$\chi^2 = 9.93^{**}$
	Low	26	16	58	
3	High	4	56	41	$\chi^2 = 5.33^{*}$
	Low	19	26	55	
4	High	7	52	41	$\chi^2 = 1.01$
	Low	13	39	48	
5	High	11	63	26	$\chi^2 = 5.46^{*}$
	Low	13	32	55	

* $p < .01$ in this direction (one-tailed test).

** $p < .0025$ in this direction (one-tailed test).

smaller than that for the betting preferences for which complete data on all *S*s were available.

RESULTS

Free Practice Trials

The results for the free practice trials in the shuffleboard game are summarized in Table 1. As can be seen in Table 1, the high n Achievement group preferred to shoot from an intermediate distance. More than half of the high n Achievement group's shots were from the intermediate zone on each of the five trials, while in only one trial were more than a third of the low n Achievement group's shots taken from this zone. The chi square tests comparing high and low n Achievement groups in frequency of shots from the intermediate vs. very easy and very difficult zones combined yield statistically significant differences in the predicted direction on four of the five trials. The low n Achievement group clearly preferred the difficult zone to either the easy or intermediate zones. Among the men high in n Achievement who did not select the intermediate zone, the difficult zone was strongly preferred to the easy zone.

In order to summarize the results for the five practice trials in one statistical test, the mean number of shots from the intermediate zone for the two motivation groups was computed. The mean number of shots from the intermediate zone (6–10 ft.) by the high n Achievement group was 2.81 and the mean for the low n Achievement group was 1.32. The difference between these means is statistically significant ($t = 2.92$, $df = 56$, $p < .005$ in the predicted direction). In addition, we can compare these means with the figure we would expect if the distribution of shots across the three zones was rectangular or random.

The chance number of shots expected from each zone on five trials would be 1.67. The mean number of shots taken by the high n Achievement group from the intermediate zone is significantly higher than 1.67 ($t = 2.63$, $df = 26$, $p < .02$), but the difference between the chance figure and the low n Achievement group mean of 1.32 is not statistically significant ($t = 1.25$, $df = 30$, $p < .25$).

Reported Expectancy of Success

There are no significant differences between the number of successful shots for the two motivation groups on either the free practice shots, when distance from the target was held constant, or on the regulated practice shots. Nevertheless the high n Achievement group reported that they expected to beat more persons than the low n Achievement group. A median test shows that 68% of the high n Achievement group ($N = 28$) expected to beat 63 or more persons in the class, while only 35% of the low n Achievement group ($N = 31$) expected to beat this many. The difference is statistically significant ($\chi^2 = 6.17$, $df = 1$, $p < .02$).

Final Test Shots

On the final test shots, the actual probabilities of hitting observed in practice and the probabilities written on the floor were such that if the theory is correct the high n Achievement group would take shots from lines closer to the target than the low n Achievement group. On the basis of the probabilities written on the floor, the region from 2 ft. to 6 ft. would be the .60 to .40 (intermediate) probability region. On the first trial of the test shots, 64% of the high n Achievement group ($N = 25$) shot from 2 ft. to 6 ft. while only 39% of the low n Achievement group ($N = 23$) shot from this region. No one shot from the one ft. line. The difference is statistically significant ($\chi^2 = 2.97$, $df = 1$, $p < .05$ in the predicted direction). Furthermore, five of the low n Achievement group but none of the high n Achievement group shot from 15 ft. on the first trial.

The two groups did not differ significantly on later trials, probably because the high n Achievement group which hit the target more often while standing closer began to move back. The percentage of high n Achievement *S*s ($N = 25$) that shot from 2–6 ft. on Trials 1–5 respectively was 64%, 52%, 48%, 44%, 48%. The percentage of low n Achievement *S*s ($N = 23$) that shot from 2–6 ft. on trials 1–5 respectively was 39%, 39%, 43%, 43%, 43%. Earlier studies (e.g., Atkinson & Reitman, 1956; Mandler & Sarason, 1952) provide sufficient basis for expecting that the high n Achievement group would perform better in a test situation than the presumedly more anxious low n Achievement group. The high n Achievement group hit the target on 30% of its shots; the low n Achievement group hit on only 16% of its shots. The difference between the mean number of hits by each group, using a log transformation to remove skewness, is statistically significant ($t = 2.42$, $df = 46$, $p < .01$ in this direction). When distance of shots is held constant, the high n Achievement group still tends to hit more often than the low n Achievement group. For example, 25.8% of 66 shots by the high n Achievement group vs. 10.7% of 56 shots by the low n Achievement group were hits from 6–10 feet. But the unequal number of shots from various distances by *S*s in the two groups makes it difficult to clearly evaluate the statistical significance of the difference holding distance constant.

Betting Preferences

In the analysis of the betting preferences, the bet most often selected was considered the preferred bet. There was a tendency on the part of all *S*s to prefer the 6/6 option, which was money already won. There was no difference between the high and low n Achievement *S*s in their willingness to accept the sure thing. In the small expected value bets 47% of the high n Achievement group ($N = 17$) and 42% of the low n Achievement group ($N = 12$) clearly preferred the 6/6 option. The chi square test of *S*s who clearly prefer 6/6 vs. those who showed any other preference does not allow rejection of the null hypothesis ($\chi^2 = .08$, $df = 1$, $p > .75$). For the large expected value options, 33% of the high n Achievement group ($N = 15$) and 41% of the low n Achievement group ($N = 22$) clearly preferred the 6/6 option. Again the null hypothesis cannot be rejected ($\chi^2 = .22$, $df = 1$, $p > .50$).

TABLE 2

PERCENTAGE OF *S*S HIGH AND LOW IN n ACHIEVEMENT SHOWING A PREFERENCE FOR INTERMEDIATE PROBABILITY BETS (4/6, 3/6, OR 2/6) OR EXTREME PROBABILITY BETS (5/6 OR 1/6) OR NO PREFERENCE IN TWO IMAGINARY DICE GAMES

Expected Value of Options	n Ach	*N*	Probability Preferences			Intermediate vs. Extremes[a]
			Intermediate (4/6, 3/6, or 2/6)	No Pref.	Extremes (5/6 or 1/6)	
\$ .30	High	17	53	18	29	$p = .04$
	Low	12	17	17	67	
\$300	High	15	33	33	33	$p = .16$
	Low	22	18	18	64	
Both groups combined	High	32	44	25	31	$\chi^2 = 7.43$
	Low	34	18	18	65	$p < .005$

[a] All significance tests are for one-tailed Fisher exact probability tests.

Since the 6/6 option is a riskless option, no prediction is made from the theoretical model about the tendency of persons to prefer it. For this reason, and since there was no difference between the two motivation groups in their tendency to prefer a sure thing, the second most preferred bet was taken as the preferred bet involving *risk* except when some bet involving risk was initially preferred even to the 6/6 option.

If an *S* preferred the 4/6, 3/6, or 2/6 bet to both 5/6 and 1/6, he was classified as preferring intermediate risk. If he preferred the 5/6 or 1/6 bet to 4/6, 3/6, and 2/6, he was classified as preferring extreme risk. If he preferred one of the intermediate bets and one of the extreme bets equally often, he was classified as showing no clearcut preference. Table 2 contains a summary of the preferences of the two motivation groups. For the small expected value bets (30¢), the predicted difference between the two motivation groups in tendency to prefer intermediate vs. extreme risk is statistically significant ($p = .04$ in the predicted direction). For the large expected value bets (\$300) the difference between the two motivation groups in tendency to prefer intermediate vs. extreme risk is in the predicted direction, but is not statistically significant ($p = .16$ in this direction). When the two conditions are combined, the chi square for the difference between the high and low n Achievement groups yields evidence of a substantial difference ($\chi^2 = 7.43, df = 1, p < .005$ in the predicted direction).

When the expected value is small (30¢), the eight *S*s in the low n Achievement group who prefer extreme probabilities are divided evenly between 5/6 and 1/6. However, when the expected value is large ($300), 10 of the 14 *S*s in the low n Achievement group who prefer extreme probabilities choose the 5/6 alternative.

A so-called "goodness of fit" test[7] shows that it is the low n Achievement group that shows a strong probability preference; the preferences of the high n Achievement group are not different from what would be expected by chance. Since three of the five bets are intermediate risk bets, .6 of the preferences should be for intermediate risk bets by chance. For the small expected value bets (30¢), the high n Achievement preferences are not different from what is expected by chance ($p = .97$), but the low n Achievement group clearly prefers the extreme risk bets more often than expected by chance ($p = .02$). For the large expected value bets ($300) the high n Achievement group's preferences are again not different from a chance level ($p = .73$), but the low n Achievement group exhibits a strong preference for extreme risk bets ($p = .003$).

Similar Risk Taking Strategies on Two Games

Table 3 contains the percentage of *S*s high and low in n Achievement who could be classified as intermediate or extreme on both the shuffleboard and betting preferences, or who showed a mixed strategy. To increase the size of the sample, the criterion of intermediacy for betting was that one of the intermediate bets was preferred as often as, or more often than, either of the extreme bets. This is a less stringent criterion of preference that approximates a median break of the preferences. For the shuffleboard game the criterion of intermediacy was that two of the five shots were taken from 6–10 ft.; again, a somewhat relaxed criterion approximating a median break. Table 3 shows that the motivation groups are quite consistent in their approach to the two games. The very significant chi square ($\chi^2 = 10.16, df = 1, p < .001$) shows that of the people exhibiting similar strategies in the two games, those preferring intermediate risk tend strongly to be high in n Achievement and those showing preference for extreme risk tend to be low in n Achievement.

TABLE 3

PERCENTAGE OF *S*S HIGH ($N = 26$) AND LOW ($N = 29$) IN n ACHIEVEMENT SHOWING SIMILAR OR DISSIMILAR STRATEGIES IN THE SHUFFLEBOARD GAME AND THE BETTING PREFERENCES

n Achievement	Strategy			Intermediate vs. Extreme
	Intermediate in Both[a]	Mixed	Extreme in Both	
High	46	42	12	$\chi^2 = 10.16$
Low	14	41	45	$p < .001$[b]

[a] Criterion for shuffleboard was 2 of the 5 shots from 6 to 10 ft; criterion for betting preferences was that an intermediate probability bet was preferred more often than or as often as an extreme probability bet.

[b] One-tailed test.

DISCUSSION

The most important feature of this study is the confirmation of the hypothesis derived from the risk taking model. Men who are high in n Achievement show a fairly strong preference for intermediate risk in a game requiring activity in which the outcome is contingent upon successful exercise of skill and a relative preference for intermediate risk in a game chance where their own skill and competen cannot control the outcome. Men who are lo in n Achievement generally prefer extreme probability alternatives in the same games. This is indeed what would be predicted from the model if we knew that in one group the motive to achieve was stronger than the motive to avoid failure, and in the other group the motive to avoid failure was stronger than the motive to achieve. However, in a college sample, the distribution of n Achievement scores probably represents a very selected group of motivated people. Thus it would seem probable that not all *S*s below the median in n Achievement have motives to avoid failure that are absolutely stronger than their motives to achieve. It need be assumed only that the avoidance motive is relatively stronger in this group than in the high n Achievement group to predict a relative difference in strength of preference for intermediate vs. extreme risk. For this reason there is little difficulty in making sense of the finding that in the initial "free" shots of the shuffleboard game the men who were high in n Achievement took significantly more than a chance number of shots from 6 to 10 ft., while the low n Achievement group took fewer than a chance number from this region, although the tendency to avoid, i.e., to depart from the chance frequency, was not statistically significant. The more surprising finding is that the low n Achievement group showed any concrete evidence of a tendency to avoid intermediate risk.

---

[7] Binomial Test (Siegel, 1956) for "goodness of fit." All probabilities are two-tailed probabilities.

As originally designed, the experiment called for a set of final test shots with objective probabilities obtained from practice trials ranging from very high to very low written on the floor. We anticipated even better results with clearly stated probabilities than in the free shots where the probability of success had to be inferred as an initial reaction to distance from the target. But the *S*s were such poor shots, and their inadequacies during practice were so obviously apparent to them, that it was necessary to make the test probabilities believable. Even so, it was necessary to exaggerate their success to get the probabilities written on the floor for the nearest lines over .50. We then made the comparison between groups that the hypothesis called for, and on the first test trial again found it confirmed even though now the high n Achievement group was physically closer to the target (i.e., in the 2–6 ft. range) than the low n Achievement group.

In the shuffleboard game, both the high and low n Achievement groups preferred the very difficult region to the very easy region. This preference can be understood in light of an additional assumption proposed earlier by Atkinson (1957, 1958a) and confirmed by other evidence in the present study: a strong motive either to approach or to avoid some goal may bias the subjective probability of the outcome in the direction consistent with the motive. In other words, the person who is high in n Achievement may have a stronger expectancy of success than is warranted by the objective conditions; and the person in whom the motive to avoid failure is very strong may consistently underestimate his chance of succeeding. The statistically significant difference between high and low n Achievement groups in estimates of the number of persons they *expected to beat* in the final shuffleboard test is clearly consistent with the assumption that there might be motivational biasing of subjective probabilities. The high n Achievement group was more expectant of success than the low n Achievement group even though up to that point in the experiment there was no difference in their level of performance of the task.[8]

If the range of distances in the shuffleboard game represents a symmetrical range of difficulty around a probability of .50, the .50 point would fall roughly in the middle, at about 7 or 8 ft. The low n Achievement group, presumably stronger in the motive to avoid failure, would bias the probability downward and their subjective .50 point would be closer, possibly at 5 or 6 ft. Clearly then, the very difficult region, 11–15 ft., would represent the most successful avoidance of the intermediate probability region. The median of the distribution of all shots by low n Achievement *S*s was at 10.92 ft. The high n Achievement group, on the other hand, should bias the probability upward, and their subjective .50 point should fall farther away, possibly at 9 or 10 ft. The point of maximum preference for the high n Achievement group should then fall in the intermediate zone, but nearer to the very difficult region, 11–15 ft., than to the very easy region, 1 to 5 ft. The median of all shots by the high n Achievement group was at 10.03 ft. In other words, the distribution of shots by the two groups shown in Table 2 is consistent with these conjectures regarding possible biasing of subjective probabilities.

---

[8] It is worth noting, at this point, the distinction to be made between operations that provide an estimate of subjective probability, e.g., how many do you expect to beat, or how many times out of 10 do you expect to hit the target, and operations that reveal the resultant motivation of the *S*, viz., the level of difficulty actually chosen or the specific goal the individual sets to *try for*. The appropriate use of these often confused operations is illustrated in the paragraphs above. An individual's estimate of the number of persons he expects to beat is used as a rough measure of his subjective probability of success; and the level of difficulty he actually chooses in shooting is used as evidence of his resultant motivation (or level of aspiration).

The tendency of relatively more anxious *S*s to avoid intermediate risk by preferring "long shots" more frequently than "sure things," here interpreted in terms of biased probabilities of success, may on certain occasions be explained in other ways. The theoretical model (see Atkinson, 1957) clearly states that extreme subjective probabilities (e.g., .10 and .90) which are equally distant from a subjective probability of .50 are equivalent in arousing weak anxiety about failure. However, the model does not necessarily imply that therefore an equal number of *S*s should avoid in each direction. Trying very difficult tasks, even when there appears to be no chance of success, may often appear more socially desirable (Edwards, 1957) than attempting the easiest tasks. Or the number of individuals who avoid failure by excessively high or low levels of aspiration may depend upon the relative frequency of persons in the sample who have developed one or the other of these strategies as a characteristic mechanism of defense. Several writers (Eysenck, 1955; Eysenck & Himmelweit, 1946; Miller, 1951) have shown that neurasthenics and hysterics are characteristically different in just this respect; the former normally set excessively high and the latter excessively low levels of aspiration. Furthermore, we should expect relatively anxious *S*s to pile up at either the very easy or very difficult ends of a risk continuum whenever the range of tasks presented is less than a symmetric range of probabilities around the subjective probability of .50. For example, in the test trials of this experiment, there were no "easy" tasks to choose. The closest lines had stated probabilities of success near .50, so the relatively more anxious *S*s avoided shooting from up close. They appeared to have higher levels of aspiration than the high n Achievement group. The result could have been just the reverse if tasks had ranged from "very easy" (i.e., very high P_s) to "moderately difficult" (i.e., P_s near .50) with no very difficult tasks to choose from.

In other words, we may expect to find n Achievement groups with higher, lower, or intermediate aspirations in relation to those of low n Achievement groups depending completely on the range of difficulty that is presented to the *S*s. It sounds very slippery indeed, to say theoretically anything can happen depending upon the conditions. Yet that is precisely what the theoretical model says.

The obvious problem for future research is to provide satisfactory independent measurement of subjective probability of success. The use of geographic distance to estimate strength of expectancy of success, although the only available measure during the free shots, may be considered the weak point of this investigation. A subsequent study by Litwin (1958) has begun to explore the systematic use of different operations for defining subjective expectancies.

In the betting preferences, the high n Achievement group preferred the intermediate vs. extreme risk bets significantly more often than the low n Achievement group when the expected monetary value was only 30¢. However, when the expected monetary value is increased to $300, the differential in risk taking preferences is reduced. Apparently an increase in the amount of money to be won—even in imagination—increases the relative strength of motivation for money qua money. In the small expected value bets, the amount of money may function as no more than a symbol of the achievement, since it is trivial in amount.

On the face of it, this evidence of a risk taking preference in gambling linked to achievement motivation is something of a paradox. Persons who are highly motivated to achieve should *not* experience any sense of accomplishment, nor anticipate it, when as a matter of fact the outcome is a chance event, one for which they cannot feel personally responsible. Nor should failure to win be reacted to as a personal threat when the motive to avoid failure is strong. The fact that the high n Achievement group *did not* show a statistically significant preference for intermediate vs. extreme risk bets when compared with the chance expectation is consistent with this supposition that games of skill and games of chance should be reacted to differently. The further washing out of the slightly greater preference for intermediate risk by the high n Achievement group when the amounts of money are increased suggests that the slight preference for intermediate risk with small amounts of money can be understood as a simple generalization of the tendency developed in competitive achievement situations requiring skill that is checked when the chance nature of the activity becomes more salient. Yet the low n Achievement group clearly avoided intermediate risk even when the amount of money was increased. This finding may mean that persons who have a long history of anxiety in competitive achievement activities fail to discriminate between games of skill and games of chance because the outcome more often seems beyond their control even in tasks requiring skill.

The present results regarding betting preferences must be considered little more than exploratory. We have already begun to follow them up. The small expected value (30¢) betting questionnaire described in this paper has been used in two small follow-up studies with male *S*s. When the results for all three studies are combined, it appears that among *S*s high in n Achievement ($N = 44$): 16% preferred 5/6, 14% preferred 4/6, 25% preferred 3/6, 7% preferred 2/6, and 11% preferred 1/6, while 27% showed no clearcut preference. Among *S*s low in n Achievement ($N = 36$): 33% preferred 5/6, 14% preferred 4/6, 11% preferred 3/6, 0% preferred 2/6, and 28% preferred 1/6, while 14% showed no clearcut preference. The chi square for intermediate (4/6, 3/6, 2/6) vs. extremes (5/6, 1/6) is 7.10 ($df = 1$, $p < .005$ in the predicted direction). In addition, "goodness of fit" tests show that the high n Achievement group's preference for intermediate vs. extreme risk is not different from what is expected by chance ($\chi^2 = .08$, $df = 1$, $p > .75$) while the low n Achievement group very strongly prefers the extreme risk bets ($\chi^2 = 12.39$, $df = 1$, $p < .001$). Preliminary data has also been collected on zero-expected value bets and on small negative-expected value bets (lose 30¢). In neither case, however, was any relationship found between n Achievement and a tendency to take intermediate vs. extreme risks.

As tenuous as these results presently appear, they are of considerable importance in light of Edwards' (1953) earlier report of probab preferences in gambling and the tendency of persons engaged in research related to probability-utility theory of decision scarcely to pay even lip service to the notion that anything but the value of money qua money affects the utility of the outcome in simple bets.

SUMMARY

This study deals with the effects of individual differences in strength of achievement motive on (*a*) goal setting inferred from choices individuals make among tasks varying in difficulty and (*b*) preferences for imaginary bets equated for expected monetary value but differing in probability of winning. The stimulus for this research is a risk taking model which predicts that, when the motive to approach success is stronger than the motive to avoid failure, tasks with intermediate probabilities of success will be preferred; but, when the motive to avoid failure is stronger, intermediate risk will be avoided and tasks with very high or very low probabilities of success will be preferred. Measures of n Achievement were obtained by content analysis of responses to the French Test of Insight. Persons with high n Achievement scores were assumed to have relatively stronger motives to approach success and those with low n Achievement scores relatively stronger motives to avoid failure.

Sixty-six college men played a shuffleboard game in which they were allowed to shoot from any of 15 lines varying in distance from the target. *S*s high in n Achievement initially shot from intermediate distances, 6 to 10 ft., significantly more often than *S*s low in n Achievement. However, when it became apparent after practice that the probability of success was .50 much closer to the target, *S*s high in n Achievement then took more shots than *S*s low in n Achievement from up close.

The high n Achievement group also preferred intermediate risk bets (4/6, 3/6, 2/6) over extreme risk bets (5/6, or 1/6) significantly more often than the low n Achievement group when the expected monetary value of the bet was small (30¢).

The results are consistent with the theory concerning motivational determinants of risk taking behavior and constitute evidence of the

possible effects of individual differences in strength of achievement motive on risk taking in games of chance as well as in tasks requiring skill.

REFERENCES

ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372. (Also reprinted in Atkinson, 1958b)

ATKINSON, J. W. Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In Atkinson, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. (a)

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. (b)

ATKINSON, J. W., & LITWIN, G. H. Achievement motive and test anxiety as motives to approach success and to avoid failure. *J. abnorm. soc. Psychol.*, 1959, **58**, 52–63.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366. (Also reprinted in Atkinson, 1958b)

BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 467–476.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EDWARDS, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.

EYSENCK, H. J., & HIMMELWEIT, H. T. An experimental study of the reactions of neurotics to experiences of success and failure. *J. gen. Psychol.*, 1946, **35**, 59–75.

FRENCH, ELIZABETH G. Development of a measure of complex motivation. In Atkinson, J. W. (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

LITWIN, G. H. Motives and expectancy as determinants of preference for degrees of risk. Unpublished honors thesis, Univer. of Michigan, 1958.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In Hunt, J. McV. (Ed.), *Personality and behavior disorders*, Vol. 1, Chap. 10. New York: Ronald, 1944.

MCCLELLAND, D. C. Some social consequences of achievement motivation. In Jones, M. R. (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln: Univer. Nebraska Press, 1955.

MCCLELLAND, D. C. Risk-taking in children with high and low need for achievement. In Atkinson, J. W. (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **16**, 115–118.

MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, **46**, 378–387.

MOULTON, R. W., RAPHELSON, A. C., KRISTOFFERSON, A. B., & ATKINSON, J. W. The achievement motive and perceptual sensitivity under two conditions of motive arousal. In Atkinson, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

RAPHELSON, A. C. The relationship between imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–18. (Also reprinted in Atkinson, 1958b)

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–326.

VITZ, P. The relation of aspiration to need achievement, fear of failure, incentives, and expectancies. Unpublished honors thesis, Univer. of Michigan, 1957.

(Received October 8, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE INTEREST VALUE AND RELEVANCE OF FEAR AROUSING COMMUNICATIONS

LEONARD BERKOWITZ AND DONALD R. COTTINGHAM
University of Wisconsin

SEVERAL years ago Janis and Feshbach (1953) tested the "common sense" belief that attitudes are readily changed by fear arousing communications. As they point out, this notion, widely shared by health and safety educators as well as many propagandists, assumes that the frightened listener becomes highly motivated to accept the reassuring recommendations advocated by the communicator. However, contrary to this popular view, Janis and Feshbach showed that opinions presented through a strong fear arousing communication were less readily adopted by high school students than opinions presented in a talk making minimal use of fear.

The Yale investigators caution against overly hasty generalizations from their findings. Conceivably, they suggest, fear appeals may prove to be highly successful under certain conditions. The present study is an attempt to specify some of these conditions.

One possibility has to do with the interest value of the communication. A communicator may not succeed in changing his audience's opinions if his message lacks sufficient interest, even though he possesses other prerequisites for successful influence, such as high credibility (Hovland, Janis, & Kelley, 1953). People required to listen to or read an uninteresting message undoubtedly suffer some frustration, particularly if the message is long. As in the case of other frustration reactions, the bored audience might respond by "withdrawing from the field," i.e., not attending to the communication, or by directing hostility toward the communicator. Thus, it is likely that relatively high interest value is necessary but not sufficient for creating opinion changes through mass communications, either because low interest produces inattentiveness or because the frustrated audience, venting its aggression, disparagingly dismisses the entire communication.

This hypothesis can easily be reconciled with the results of the earlier study. There, even though the strong appeal was of somewhat greater interest to the *S*s than the moderate and minimal fear appeals (though the former condition was least effective in changing attitudes), the majority of *S*s in all conditions thought that "all or most" of the communication they received was interesting. The message, then, probably was above threshold in interest for most *S*s in all conditions, permitting the operation of other attitude changing variables. However, other communications in other settings may be low in intrinsic interest for the audience. The dramatic nature of anxiety provoking material may heighten the interest value of a communication above the required threshold so that under certain conditions the strong fear appeal may successfully change opinions. In contrast, a minimal fear appeal, because less interesting, may produce little opinion change.

A second proposition concerning the success of a fear appeal assumes individual differences in the strength of emotional reactions to the communication. The importance of this assumption can be seen in the theoretical analyses made by Janis and Feshbach. To account for their results they raise the possibility of "defensive avoidance". "When fear is strongly aroused but is not fully relieved by the reassurances contained in a mass communication, the audience will become motivated to ignore or to minimize the importance of the threat" (1953, p. 90). In other words, the audience defensively avoids accepting the beliefs advocated by the communicator. However, in a later re-examination of their data (1954), the Yale psychologists also point out that some *S*s will become more fearful than others as a result of the threatening appeal. Contending that people with a chronically high anxiety level are relatively vulnerable to fear arousal, they show that "persons who manifest symptoms of acute anxiety in their everyday lives" were consistently less influenced by the strong appeal than low anxiety *S*s. The former *S*s, presumably because of their stronger fear, exhibited a stronger defensive-avoidance reaction.

In this case, then, the degree of fear aroused by the threatening communication seems to be a direct function of *S*s' characteristic level of anxiety. Extrapolating from Miller's (1948) conflict model relating the strength of the individual's avoidance tendencies to his psychological distance from a desired yet feared goal, we can also propose that the magnitude of the avoidance-fear reaction increases with the relevance of the message to the needs of the audience members. Those *S*s for whom the message is highly relevant obviously are closer to making the appropriate goal responses called for by the communication, i.e., they are closer to using the advocated beliefs than the *S*s for whom the material is less relevant. Therefore, since *S*'s fear and the resulting tendency-to-avoid response increases with his approach to the feared situation in which the response would be elicited, a stronger defensive avoidance tendency should be produced by a relevant than by an irrelevant communication.

Combining the two lines of reasoning presented above, we predict: (*a*) To the extent that the minimal fear appeal is relatively uninteresting to the audience, there should be little, if any, opinion change in the *S*s receiving this communication. (*b*) Assuming that the strong appeal is of sufficient interest, the less relevant the material to the *S*s in this condition the greater the opinion change in the direction advocated.

METHOD

Procedure

Two separate experiments were conducted employing different *E*s and different "communicators" but using the same communications in the same procedure.[1] In each case, the *S*s, all volunteers from introductory psychology courses, were recruited ostensibly as part of a testing program designed to evaluate the effectiveness of various lecture techniques. When they came to the assigned rooms, in groups of at least 18 members, the *S*s in the experimental conditions listened to a tape recorded lecture on automobile safety belts, and then answered a brief questionnaire indicating their opinions of the lecture, the speaker, and the desirability of having safety belts. A control group responded only to the safety belt items without hearing the tape recording.

Two lecture conditions were employed in each study: minimal and strong fear arousal. In the former condition, after a 370-word introduction which was common to both lectures, the speaker went on to give a 350-word emotionally neutral argument elaborating the advantages of automobile safety belts and relying mainly upon statistical data together with one cartoon-like slide. This special material, read by the same speaker, not only was longer in the strong fear condition (510 words) but also was more personal and dramatic, and made use of 10 slides, most of which depicted gruesome car accidents. Essentially the same arguments were used in both lectures. The lectures remained unchanged from one study to the next, although a different male speaker was employed in each experiment.

*S*s' attitudes towards safety belts, as assessed by an 18-item Likert scale included in a questionnaire administered immediately after the experimental lecture (or the day following the experimental sessions but without a lecture in the case of the control *S*s) were then compared with their attitude scores on the same items approximately one month earlier. For this pretest the *S*s, who had volunteered to serve in a study of "the psychology of normal people", had been brought together in relatively large groups and given a 123-item questionnaire dealing with such topics as national tariffs, movies, Jews, and the punishment of criminals, as well as automobile safety belts. The *S*s were told that *E* was interested in determining how college students felt about these issues. Several weeks later a different *E* contacted each of the *S*s, informed him (or her) of the lecture effectiveness study, and solicited his participation in the experiment. No mention was made of the earlier attitude testing.

Questionnaire Measures

Two items in the postexperiment questionnaire were designed to test the effect of the fear arousal manipulations. The first of these items reads, "I would classify the presentation as . . . " with the responses ranging in five steps from "very pleasant" through "pleasant," "nonstimulative," and "unpleasant" to "very unpleasant." The second item was worded, "The presentation of the material was . . . ," with alternatives "*a*) boring, *b*) of indifferent quality, *c*) thought provoking and *d*) disturbing."

The responses to the items were combined in Experiment I by weighting each alternative in terms of the amount of discomfort indicated and adding the two scores. The higher the score, then, the higher the admitted tension evoked by the communication. However, in Experiment II the items were kept separate for two reasons. First, answers to the items exhibited a somewhat different pattern in the second study. Fewer people in the latter investigation described themselves as very upset by the lecture, particularly in response to the second question, while a greater proportion of the *S*s indicated the material was uninteresting to them. Second, the greater frequency of responses indicative of boredom in Experiment II made it possible to test the effects of communication interest-value upon opinion change.

---

[1] Experiment I was carried out as part of a course in experimental social psychology conducted by the senior author. Each member of the class participated in the research in some way. Experiment II was carried out by D. R. Cottingham under the supervision of L. Berkowitz.

The relevance of the lectures to the *S*s also was assessed somewhat differently in the two investigations. In the first experiment only two levels of relevance were established, defined solely in terms of frequency of using a car: once a week or less vs. at least several times a week. Car ownership as well as frequency of car usage was taken into account in determining three levels of relevance in Experiment II. Those who indicated that they both owned an automobile and drove it at least several times a week were placed in the High Relevance group. The Medium Relevance group either owned a car and drove it less frequently, or did not own an automobile but rode in one at least several times a week. The Low Relevance *S*s neither owned a car nor rode in one more often than on weekends or once a week.

The scale to measure attitude towards safety belts included statements ranging from "automobile safety belts should be required by law" to items maintaining that their value had not yet been proven. In both experiments the *S*s indicated whether or not they agreed with the item. For convenience, we will refer to a change from the initial attitude score to the post-experiment score as a change in the number of items accepted, although change in the number of negative items rejected is also involved. The mean change in each condition was toward the advocated beliefs.

To summarize, the experimental procedure was the same in Experiments I and II, although a different speaker was used in each. The major difference between the two studies lies in the classification of relevance levels within the strong, minimal, and control conditions. The *S*s in Experiment I were divided into two relevance levels in terms of their frequency of car usage, while the *S*s in the second study were divided into three such levels on the basis of both car ownership and usage. The main dependent variable, a Likert scale of attitudes towards automobile safety belts, was the same in both investigations, but there are some differences in the amount of supplementary information available. Where Experiment I affords only a single composite index of discomfort induced by the communication, Experiment II provides information concerning both the tension felt by the *S*s and their degree of interest in the communication.

RESULTS

Experiment I

Six groups of 13 *S*s each were formed by dividing each of the experimental conditions (the strong and minimal fear appeals conditions) and the control condition into the two relevance levels. Table 1 shows the mean discomfort score induced by the communications in each of the resulting four experimental groups. It is clear that the manipulation of fear arousal was successful. The *S*s in the strong fear arousal condition admitted to significantly greater feelings of tension than the *S*s receiving the minimal fear appeal ($t=2.51$; $p=.02$).[2]

TABLE 1
MEAN DISCOMFORT SCORE IN EACH OF THE EXPERIMENTAL GROUPS
(Experiment I)

Condition	Mean Dis-comfort Score	Significance of Differences
Strong Fear Arousal		
High Relevance	6.54	H.R. vs. L.R.:
Low Relevance	5.62	p = .05
Minimal Fear Arousal		
High Relevance	4.92	H.R. vs. L.R.:
Low Relevance	5.08	NS

Note.—Strong vs. Minimal: p = .02.

The next table presents the mean opinion change in the direction advocated by the communication within each of the six conditions. As is readily apparent, there are no reliable differences among the experimental and control conditions when all the *S*s are combined across relevance levels. However, when we consider the relevance of the message, our theoretical expectations are confirmed. The Low Relevance *S*s in the strong fear condition exhibited a significantly greater change in the direction advocated than either the High Relevance *S*s in this condition or the similar Low Relevance *S*s in the control condition. (These differences, the only ones reliable, are significant at $p < .05$.)

The table also suggests that the strong fear appeal was somewhat (but not significantly) more effective than the minimal fear appeal for the Low Relevance *S*s. It may be that the former communication possessed greater interest value for the audience, permitting the members to be convinced by the logic of the speaker's message, while the same arguments could not be effective when they were embedded in an uninteresting communication.[3]

---

[2] The error terms in the two-tailed t tests used for the condition comparisons in this paper are based upon the residual mean squares obtained from preliminary analyses of variance.

[3] The relatively great increase in favorableness of attitudes towards safety belts in the control conditions suggests that many *S*s came to believe automobile safety belts were desirable after being stimulated to think about this issue.

TABLE 2

MEAN CHANGE IN ATTITUDE TOWARD SAFETY BELTS: EXPERIMENT I
(N = 13 in each condition)

	Control	Minimal Fear	Strong Fear
High Relevance			
Initial attitude	4.19	3.81	4.04
Mean change	3.29	3.48	3.14
Low Relevance			
Initial attitude	4.14	4.34	4.40
Mean change	2.90	3.83	4.25

TABLE 3

NUMBER OF *S*S GIVING THE FOLLOWING REACTIONS TO THE ITEM, "I WOULD CLASSIFY THE PRESENTATION AS . . .", IN EXPERIMENT II

Condition	Pleasant	Non-stimulative	Un-pleasant
Strong fear appeal	16	2	8
Minimal fear appeal	7	12	1

Note.—Chi square (employing Yates' correction) = 14.26; $p < .001$.

Experiment II

The second experiment, using 64 *S*s, was designed to test the foregoing interpretations. One item on the postquestionnaire, identical to that employed in the earlier study, indicates that the strong fear appeal again created a somewhat stronger feeling of discomfort in the audience than the minimal fear appeal. The data are presented in Table 3. The speaker employed in this study apparently was less effective than the speaker used in the earlier investigation; although the script was identical, only about 31% of the *S*s in the strong fear condition (as compared with only one *S*—5%—in the minimal fear condition) now admitted to some feelings of unease. The chi square for this difference, based upon only those *S*s classifying the presentation as "unpleasant" and employing Yates' correction, is 4.00 ($p = .05$).

The table also indicates that the people in the strong fear condition were more likely to call the presentation "pleasant," and were less likely to be "indifferent" to it, than the *S*s in the minimal fear condition. The chi square for the former reaction is 3.52 ($p < .10$), and for the latter is 7.14 ($p < .01$). Taken together, these differences suggest that the strong appeal was more interesting to the audience than the minimal fear arousing lecture. Further evidence is provided in the analysis of the second questionnaire item summarized in Table 4. Forty per cent of the *S*s in the strong fear condition indicated on this item that the lecture had low interest value, while this figure rose to 75% in the minimal fear condition.

According to the present hypotheses, then, opinion changes are more likely to occur through the relatively interesting strong fear arousing lecture than in the more boring minimal fear arousing condition. This expectation is confirmed by the data presented in Table 5.

As can be seen in Table 5, the over all pattern of results is similar to the findings obtained in Experiment I. First, the *S*s receiving the fairly uninteresting minimal fear appeal were not more likely than the control *S*s, to change in the direction advocated, and there was no significant relationship between the relevance of the communication and the amount of opinion change for the *S*s in these conditions. Second, consistent with Prediction *b*, in the case of those *S*s receiving the more interesting strong fear-arousing communication, the less relevant the material the greater the opinion change in the direction advocated. In the strong fear condition the Low Relevance *S*s have a greater mean change score (in terms of number of items accepted) than either the High or Medium Relevance *S*s (though the latter t does not attain significance; t with the High Relevance *S*s is 3.23; $p = .01$), while the Medium Relevance *S*s exhibit a reliably greater change than the High Relevance *S*s ($t = 2.28$; $p = 0.05$). Moreover, the Low Relevance *S*s receiving the strong fear appeal have a significantly greater mean change score than any of the

TABLE 4

NUMBER OF *S*S CLASSIFYING THE EXPERIMENT II PRESENTATION AS LOW IN INTEREST-VALUE ("BORING" OR "OF INDIFFERENT QUALITY") OR NOT LOW IN INTEREST-VALUE ("THOUGHT PROVOKING")

Condition	Low Interest	Not Low Interest
Strong fear appeal	10	15
Minimal fear appeal	15	5

Note.—Chi square = 5.99; $p = .05$.

TABLE 5

MEAN CHANGE IN ATTITUDE TOWARD SAFETY BELTS: EXPERIMENT II

	Control	Minimal Fear	Strong Fear
High Relevance			
N	6	7	11
Initial attitude	6.0	6.4	6.3
Change score	2.2	1.7	2.7
Medium Relevance			
N	6	6	9
Initial attitude	5.7	5.5	7.2
Change score	1.2	3.3	6.2
Low Relevance			
N	6	7	6
Initial attitude	7.3	6.7	6.8
Change score	2.2	2.7	8.3

groups in the minimal fear and control conditions ($p < .01$ in every case with the exception of the .05 level difference with the Medium Relevance *S*s in the minimal fear condition). None of the other differences is significant.

Other Effects of Fear Arousal

In keeping with the Yale approach to studies of opinion change (Hovland, Janis & Kelley, 1953), Janis and Feshbach asked whether their condition differences could be attributed to differences in the learning of the material presented and/or differences in *S*s' attitudes toward the communicator. According to the first of these possibilities, the defensive avoidance tendency provoked by the strong fear arousal could have produced inattentiveness to the communication. However, they obtained no evidence for such a process; *S*s in the strong fear condition learned as much of the communicated information as those in the less fearful conditions. The present theoretical formulation also necessitates obtaining indices of the degree of attention in the various communication conditions. We have proposed that the frustrating experience of listening to a boring speech may result in greater inattentiveness to the lecture. Again, however, there is no evidence of differences in the learning of the lecture material. On an information test included in the postexperiment questionnaire (in Experiment II), *S*s in both experimental conditions at each relevance level were correct, on the average, on approximately 7 of the 11 items.

The second possibility mentioned and then discarded by Janis and Feshbach was that the fear-arousing communication provoked hostility toward the communicator which in turn interfered with the acceptance of his argument. We also have suggested that aggression towards the communicator may be a factor in producing little opinion change, but that frustration may develop during the relatively uninteresting minimal fear arousing lecture. To test this, an item in the postexperiment questionnaire used in the second study asked *S* to indicate whether the lecturer sounded very pleasant, pleasant, neutral, unpleasant, or very unpleasant. We assume that *S*s who feel angry towards the speaker would be more likely to judge him as unpleasant. Contrary to our expectations, 31% of the 26 *S*s in the strong fear condition rated the speaker as unpleasant, while only 10% of the 20 *S*s in the minimal fear condition made this judgment. However, the chi square value for this difference is only 1.69, employing Yates' correction, which is nonsignificant.

It would appear, then, that the relative ineffectiveness of the minimal fear arousing lecture cannot be traced either to failure to attend to the communication or aggression toward the communicator. Any hostility instigated by the presumably frustrating experience of listening to the uninteresting lecture probably was directed at the communication (as indicated by the disparaging evaluations of the presentation in the minimal fear condition) and/or to the experimenters.

DISCUSSION

Little direct evidence has been presented so far for the process hypothesized to account for the relatively great opinion change in the Low Relevance subgroup. Contrary to the present formulation, these *S*s within the strong appeal condition may have accepted the communication, not because they had a low avoidance tendency, but because, lacking driving experience, they were more susceptible to authoritative persuasion in this area. Clearly, they were not generally more susceptible to arguments favoring safety belts; the Low Relevance *S*s in the minimal appeal condition did not tend to adopt the advocated beliefs. However, as we already have suggested, the low level of opinion change under minimal fear appeal may have been the result of an uninteresting communication, while the strong appeal may have been more convincing. Given this necessary interest value, the inexperienced Low Relevance *S*s could have been more susceptible to an expert's persuasion.

Nevertheless, there is reason to believe that the different relevance groups were quite similar in their driving experience. With the exception of two people at each relevance level, all *S*s in the second experiment indicated, in response to a questionnaire item, that they had been driving more than three years. The *S*s, it may be noted, were on the average between 19 and 20 years old.

But what is really needed to buttress the hypothesis that relevance increases the defensive-avoidance reaction to fear-arousing communications is evidence that anxiety (or discomfort) varies directly with relevance. Some of the data suggest that this is indeed the case. The results for Experiment I, presented in Table 1, indicate that the High Relevance *S*s receiving the strong fear appeal were significantly more upset by the communication than the *S*s for whom the message was less pertinent, while there was no difference in admitted tension between the two relevance groups in the minimal fear condition. This significant difference in the strong fear condition was not obtained in Experiment II because of the small numbers of cases, but here too there are indications of greater anxiety in the more relevant groups. Findings reported earlier suggest that evaluations of the lecturer in this study tend to bear a direct relationship to the discomfort created by his communication; he was more likely to be judged as unpleasant in the strong appeal than in the minimal appeal conditions. Closer inspection of the data reveals that the lowest ratings of the speaker were given by the High and Medium Relevance *S*s within the strong appeal treatment. Three of the former called him unpleasant, five in the latter subgroup did so, as compared with none of the Low Relevance *S*s. Employing Fisher's exact probability test, the difference between the High and Low Relevance *S*s is nonsignificant ($p = .24$), but there is a significant difference ($p = .04$) between the Medium and Low Relevance *S*s. Apparently, the fear-arousing speaker was somewhat more unpleasant to the High and Medium Relevance *S*s than to the Low Relevance *S*s.

In both experiments, then, there are indications that relevant fear-arousing communications produce greater tension in the audience than less relevant messages. If this should be the case, the present results are consistent with an hypothesis advanced by Janis and Feshbach (1954), who propose that a fear appeal is especially likely to be accepted when a relatively weak anxiety drive is aroused. Further research is necessary, however, before we can conclude with certainty that the High Relevance *S*s are less likely to accept the fear arousing appeal because of a tendency toward defensive avoidance.

SUMMARY

An experiment by Janis and Feshbach (1953) indicated that opinion change in the direction advocated by a communication is less likely to occur when the communication arouses strong fear than when it arouses less fear. In explanation, they suggested that fear arousal caused the *S*s defensively to avoid accepting the communicator's conclusions. The two experiments reported here were designed to show that a strong fear appeal could be more convincing than a weak one when (*a*) the communication is low in interest value and the dramatic nature of the "strong" communication makes it considerably more interesting than the "weak" communication, and (*b*) the communication is of low relevance to the actions of the audience.

The two experiments, employing college students who listened either under strong fear or weak fear conditions, to a tape recorded lecture advocating the use of automobile safety belts, agreed in confirming these expectations. There was little opinion change with the relatively uninteresting minimal fear lecture, while the degree of opinion change produced by the more interesting strong fear lecture was inversely related to the relevance of the material to the *S*s. Miller's hypothesis concerning the slope of the avoidance gradient was cited in accounting for the findings with the relevance dimension.

REFERENCES

HOVLAND, C., JANIS, I., & KELLEY, H., *Communication and persuasion*, New Haven: Yale Univer. Press, 1953.

JANIS, I., & FESHBACH, S. Effects of fear-arousing communications. *J. abnorm. soc. Psychol.*, 1953, **48**, 78–92.

JANIS, I., & FESHBACH, S. Personality differences associated with responsiveness to fear-arousing communications. *J. Pers.*, 1954, **23**, 154–166.

MILLER, N. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

(Received October 15, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

EVALUATIONAL REACTIONS TO SPOKEN LANGUAGES[1]

W. E. LAMBERT, R. C. HODGSON,[2] R. C. GARDNER, AND S. FILLENBAUM[3]
McGill University

SPOKEN language is an identifying feature of members of a national or cultural group and any listener's attitude toward members of a particular group should generalize to the language they use. From this viewpoint, evaluational reactions to a spoken language should be similar to those prompted by interaction with individuals who are perceived as members of the group that uses it, but because the use of the language is one aspect of behavior common to a variety of individuals, hearing the language is likely to arouse mainly generalized or stereotyped characteristics of the group. Thus, when one hears a radio broadcast of an international meeting and encounters passages of a foreign language, one's evaluational reactions to the communication are attributable, in part, to the language used and likely reflect generalized attitudinal reactions to the group that uses it.

The purpose of the present investigation was to determine the significance spoken language has for listeners by analyzing their evaluational reactions to English and French. Since we were interested in reactions that are attributable primarily to the language itself, we attempted to minimize the effects of both the voice of the speaker and his message by employing bilingual speakers reading the same message in two languages. In view of previous studies (see Licklider & Miller, 1951, p. 1070f) which have shown that evaluations of personality based solely upon voice have little or no reliability, it was predicted that the differences in the favorableness of any *S*'s evaluations of the French and English guises of speakers would reflect his attitude toward members of his own and members of the other language group. The study was carried out with *S*s living in Montreal, a community whose history centers largely in a French-English schism which is perhaps as socially significant for residents of the Province of Quebec as that between the North and the South is for Southerners in the United States.

METHOD

Procedure

A 2½ min. passage of French prose of a philosophical nature was translated into fluent English and tape recordings were made of the voices of four male bilinguals each of whom read both French and English versions of the passage. Recordings were also made of the voices of two other men, one reading the passage in English, the other in French. There were, then, 10 taped voices, four of which were "matched," each speaker using both languages, and two used as "filler" voices and for practice. The 10 voices were presented to *S*s in alternating French-English order starting with the two filler voices and allowing the maximum possible interval between successive presentations of the English and French guises of any speaker. Evaluational reactions to the matched voices only were examined.

The study was introduced as an experimental investigation of the extent to which peoples' judgments about a speaker are determined by his voice and *S*s were reminded of the common phenomenon of having a voice on the radio or telephone summon up a picture in the listener's mind of the person speaking. *S*s were *not* told that they were going to hear some of the voices twice, but rather that they would hear 10 recorded male voices, all reading the same passage, five in French and five in English. The two languages, it was said, were being used to give greater scope to the experiment. In this regard, *S*s were given copies of the French and English versions of the text they were to hear so that they could be acquainted with the message and the languages and be better prepared to pay attention only to the voices of the speakers. There was no indication that any *S* became aware of the fact that bilingual speakers were used.

The *S*s were also given a response sheet for each voice which directed them to rate each of 14 traits on 6 point scales ranging from "very little" (*fort peu*, for French *S*s) to "very much" (*beaucoup*). The traits used were: height (*taille*), good looks (*attrait physique*), leadership (*apte à diriger*), sense of humour (*sens de l'humour*), intelligence (*intelligence*), religiousness (*pieux*), self-confidence (*confiance en soi*), dependability (*digne de confiance*)[4], entertainingness (*jovialité*), kindness (*bonté*), ambition (*ambition*), sociability (*sociabilité*), character (*caractère*) and general likeability (*est-ce qu'il est sympathique?*). Each voice was played once and *S*s rated that voice on the scales both

---

[1] This research was supported by the Canadian Defence Research Board, Grant Number D77-94-01-10, and by a Ford Foundation grant to D. O. Hebb. We are grateful to Henri Barik for his computational assistance.

[2] Now with the Aluminum Company of Canada.

[3] Now at the University of North Carolina.

[4] *Digne de confiance* translates to both "dependable" and "trustworthy."

while it was playing and during the 90 sec. interval between voices. At the bottom of each response sheet *S*s were asked to indicate in what occupation this man would likely be found.

Attached to the back of the 10 response sheets, facing down and covered by a blank sheet, were several questionnaires to be completed. This packaging of questionnaires and response sheets discouraged *S*s from looking ahead in the booklet and also permitted them to answer all items anonymously. *S*s were asked only their age, place of birth and religious affiliation. After the ratings were obtained on all voices, *S*s were directed to turn to the first questionnaire.

Importance ranks of traits. The list of traits used to describe each voice was presented again and *S*s were asked to rank the traits in terms of their desirableness in friends. It was felt that this information would offer a more personally relevant set of traits for analysis.

Prejudice scale. In view of its wide use as an index of generalized prejudice and ethnocentrism, we incorporated Forms 40 and 45 of the California F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) as one of the independent measures of attitude toward outgroups.

Attitude scales. *S*s were asked to complete 14 incomplete sentences designed to elicit attitudes toward both their own and the other language group. Examples are: English Canadians think . . .; Children of French Canadian parents . . . ; The more I get to know French Canadians . . . ; French Canadians . . . ; English Canadians . . . ; and The more I get to know English Canadians. . . . The items were translated into French for the French speaking *S*s. Two separate scores were assigned each *S* on the basis of his completions: an "attitude toward English group" score and an "attitude toward French group" score each comprising the number of favorable statements made minus the number of unfavorable statements made about each group. Two judges independently scored the items for each sample and the interjudge reliabilities are, for the English sample, $r = .69$ and $r = .86$ for attitudes toward English and French respectively and for the French sample $r = .75$, and $r = .80$.

Preference scale. *S*s indicated their preference for English or French Canadians as: marital partners, friends, neighbours, tenants, colleagues, and political candidates, by ranking which they would rather have for each case. Preference scores were the sum of differences in ranks given to French and English Canadians for all items.

Degree of bilingualism. We reasoned that differences in evaluational reactions to the English and French guises would be less pronounced for those *S*s who had experience with both languages and presumably both linguistic communities as well. *S*s checked their degree of facility in the use of the other group's language by indicating how well they could speak, read, and write, and whether they could do so: not at all, a little, fairly well, or fluently. An index of bilingualism was computed for each *S* by assigning zero to "not at all" entries, and 3 to "fluently." We doubled the score for ability to speak the language on the assumption that speaking experience was most important for the purpose of the study. A maximum score would be 12. A bimodal distribution of scores was obtained from both samples and both were split at the same point so that *S*s with scores 0 to 7 are categorized as being less bilingual and those with scores 8 to 12 as being more bilingual.

Subjects

The English speaking sample comprised 64 students, taking the first course in Psychology at McGill University, who volunteered to participate. The average age for the group was 18.8 years and both sexes were approximately equally represented.[5] The introduction and procedures were presented to the group by the first two authors in English. All *S*s used English as their school and primary social language, and all but nine who were born in Europe learned English as their first language.

The French speaking sample was made up of 66 male students in their final year at a classical French *collège* in Montreal who were at approximately the same educational level as the English sample, with an average age for the group of 18.2 years. The study was presented to the students by a professor of the *collège* and the first author, in French. The experiment took the place of a regular lecture. All but one of these *S*s were born in Canada and all but four in the province of Quebec. All used French as their home, school, and primary social language. The faculty and students speak grammatically correct French, of course, but in general they have a distinctive French Canadian accent.

Speakers

Much care was taken to select appropriate speakers who could take on both French and English guises (in the sense of habitual manners of speaking). We chose from acquaintances in the English speaking community three bilinguals (Bla, Cou, and Leo) who spoke faultless English and yet were trained in French schools in Canada through the graduate level and used French in their homes, or, in the case of Leo, in social or work situations. Several recordings had to be made of each speaker to have errorless readings and in the case of Bla to perfect the pronunciation of a few English words. The final recordings of all four speakers were judged to be spoken in perfect French and English by three bilingual judges. The judges agreed that speakers Bla and Cou spoke with French Canadian accents and that Leo spoke with a marked French Canadian accent characteristic of those who work "in the bush" where, in fact, his style of pronunciation was acquired. The

---

[5] Thirty-two *S*s indicated that their religion was Jewish, and, to determine their comparability with the rest of the sample, a separate analysis was carried out for Jewish and non-Jewish *S*s. For both subgroups, correlations were calculated between each of the four attitude measures and the comparative favorableness of evaluations of the English and French guises considering all traits as well as those most desirable for friends. The comparison of correlations indicated that there were no statistically significant differences between the two subgroups at any point and therefore the Jewish and non-Jewish *S*s were combined.

TABLE 1

t VALUES FOR SIGNIFICANCE OF DIFFERENCES IN EVALUATIONS OF ENGLISH AND FRENCH GUISES OF SPEAKERS

Trait	English *S*s' Judgment of Speakers					French *S*s' Judgments of Speakers				
	Cou	Bla	Leo	Tri	Over-all	Cou	Bla	Leo	Tri	Over-all
Height	8.63**	4.67**	4.05**	2.59*	8.83**	5.50**	2.32*	2.41*	1.04	4.58**
Good Looks	7.16**	4.10**	2.66**	−.90	5.78**	7.50**	6.00**	7.26**	1.73	9.77**
Leadership	1.72	.69	1.41	−1.82	1.20	4.05**	9.32**	12.10**	.82	11.06**
Sense of Humor	−.58	−.14	2.25*	−3.48**	−2.16*	2.67**	.58	.64	−.30	1.40
Intelligence	1.44	.11	2.25*	1.98*	2.40*	4.54**	9.17**	11.73**	1.00	10.22**
Religiousness	−.67	−.33	−.26	1.81	.27	−1.79	−3.11**	−1.60	−1.14	−2.94**
Self-confidence	−.50	−.47	1.13	−1.33	−.69	5.37**	7.56**	10.00**	−2.00*	8.28**
Dependability	1.61	−.15	3.51**	2.74**	3.11**	2.89**	4.36**	7.50**	1.89	6.44**
Entertainingness	1.77	1.67	.00	−4.44**	−.84	.44	−.27	−.81	.22	−.21
Kindness	2.40*	2.47*	1.67	.74	3.37**	.10	−4.12**	−2.65*	−.74	−2.96**
Ambition	.45	2.43*	2.33*	.84	2.83**	3.06**	4.64**	7.53**	−1.53	5.89**
Sociability	.24	−.16	−.17	−4.13**	−1.44	3.44**	2.10*	3.47**	.29	3.72**
Character	3.07**	1.00	3.17**	.12	3.08**	3.95**	6.81**	8.68**	.31	8.51**
Likability	2.38*	.26	1.38	−1.21	1.22	1.83	.80	2.05*	.79	2.00*

Note.—Positive entries indicate that English guises are evaluated more favorably than French, and minus entries indicate a more favorable evaluation for French guises. This is so for both French and English *S*s' judgments.

* Significant at .05 level.

** Significant at .01 level, two-tailed tests.

fourth speaker, Tri, spoke French with an accent that was judged as indistinguishable from that used in France. It was felt that these variations in accent and style would not only make the task more interesting for *S*s but would also lead to new aspects of the problem for further study, e.g., the reactions of both samples to Parisian French (Tri) and to caricatured French Canadian French (Leo). However, since the present study was not designed to deal comprehensively with variations in accent, attention will be mainly given to over-all comparisons of French and English languages until the final section where accent differences will be discussed.

RESULTS

Favorableness of reactions: English Ss' *judgements*. For each *S* on each of the 14 traits the difference (D) between the evaluations of the English and French guises of each speaker was noted, with account taken of the direction of the difference. These D values were then summed over the 4 speakers and over *S*s and the departure of the D scores from zero was examined. As can be seen in the over-all column of Table 1, the English *S*s evaluated the following seven traits significantly more favorably for the English than for the French guises: Height, Good looks, Intelligence, Dependability, Kindness, Ambition, and Character.[6] There were no significant differences in evaluations of Leadership, Religiousness, Self-confidence, Entertainingness, Sociability, and Likability. The French guises were evaluated significantly more favorably on Sense of Humor, but this finding is not clearcut since for one speaker the English guise was evaluated significantly more favorably on this trait. There was considerable consistency in the data for three of the four voices but the fourth (Tri) differed from the others on a number of traits.

*S*s had ranked the 14 traits for their desirability in friends. For each *S* two ΣD scores were obtained, one for the three traits judged most desirable and one for the three traits judged least desirable, by noting the difference in evaluations of the English and French guises for each set of traits. English *S*s evaluated the English guises significantly more favorably both for desirable traits ($t_{55df} = 2.13$) and the less desirable traits ($t_{53df} = 4.33$). There was no difference between the most and least desirable traits with regard to this preference ($t_{49df} = 1.72$).

Favorableness of reactions: French Ss' *judgments*. A similar analysis was carried out for the French *S*s. They evaluated the English guises of the speakers significantly *more* favorably than the French guises for the following 10 traits: Height, Good looks, Leader-

---

[6] Although a rating toward the upper end of the scale does not necessarily indicate favorableness in the cases of "height" or "ambition," the term "favorableness" is a meaningful description for the majority of the judgments.

TABLE 2

t VALUES FOR SIGNIFICANCE OF MEAN DIFFERENCES IN EVALUATIONS, ENGLISH VS. FRENCH JUDGES

Trait	English Guises	French Guises
Height	3.76**	0.30
Good Looks	1.78	4.59**
Leadership	−2.55*	6.19**
Sense of Humor	0.62	3.10**
Intelligence	−2.87**	2.82**
Religiousness	2.44*	−0.15
Self-confidence	−3.02**	4.85**
Dependability	−1.32	1.97*
Entertainingness	−0.52	−0.42
Kindness	1.82	−3.18**
Ambition	−0.57	2.28*
Sociability	−1.87	2.14*
Character	−1.22	3.60**
Likability	−1.34	−0.31

Note.—Positive entries indicate that English judges evaluated a particular trait more favorably on the average than did French judges; minus entries indicate more favorable evaluation for French than English judges.

* indicates a t is significant at the .05 level, on two-tailed tests.

** indicates a t is significant at the .01 level, on two-tailed tests.

ship, Intelligence, Self-confidence, Dependability, Ambition, Sociability, Character, and Likability (Table 1). There was no difference in evaluations with regard to Sense of Humor and Entertainingness, and the French guises were evaluated significantly more favorably on Religiousness and Kindness. The results for three of the four speakers were consistent and again speaker Tri was the exception. The English guises were evaluated significantly more favorably both for the three most desirable traits ($t_{37} = 7.17$) and for the three least desirable traits ($t = 7.58$) and, as in the case of the English *S*s, there was no difference between the most and least desirable traits ($t < 1$).

Favorableness of reactions: English Ss' *vs. French* Ss' *judgments.* The evaluations of the English guises given by the English judges were compared with the evaluations of the English guises given by the French judges, and a similar analysis was carried out for the evaluations of the French guises by the two groups of *S*s. Table 2 shows that for the English guises the English *S*s evaluated Height and Religiousness significantly more favorably than did the French *S*s. There were no differences between English and French *S*s with regard to: Good Looks, Sense of Humor, Dependability, Entertainingness, Kindness, Ambition, Sociability, Character, and Likability, and the French *S*s evaluated the English guises significantly more favorably than did the English *S*s on Leadership, Intelligence, and Self-confidence. The comparisons with respect to the French guises are also given in Table 2. It can be seen that the English *S*s evaluated the French guises significantly *more* favorably than did the French *S*s on the following nine traits: Good looks, Leadership, Sense of Humor, Intelligence, Self-confidence, Dependability, Ambition, Sociability, and Character. There were no differences in evaluations for: Height, Religion, Entertainingness, and Likability, and the French *S*s evaluated the French guises significantly more favorably than did the English *S*s on Kindness.

Attitudinal measures. The correlations between the various measures of attitude and the comparative favorableness of evaluations of guises are given in Table 3 for the English *S*s. It is striking that none of the attitude measures correlate with the degree of favorableness of reactions to English and French guises, with the exception of attitudes to their own group as reflected in responses to the incomplete sentences, a measure that just meets significance requirements.

Although the pattern of correlations is also low for the French *S*s, the comparative favorableness of evaluations of guises for this sample is somewhat more clearly related to both attitudes toward and preference for own group. It should be noted that the French *S*s on the average clearly showed more favorableness to English than French guises, in fact only 14% of the *S*s showed the expected trend of more favorable reactions to French over English versions. This was not the case for either the preference or attitude measures where the majority of *S*s (85% and 62%, respectively) preferred and showed more favorable attitudes toward the French over the English group. The significant correlations in Table 4 therefore indicate that the more the French *S*s show preference for and favorable attitudes toward their own group the less they overrate English guises.

The relation of bilingualism to other variables. *S*s categorized as having a high degree of bilingualism were compared with those having little experience or skill in the other group's language on all attitudinal variables as well

TABLE 3

RELATION OF ATTITUDINAL MEASURES TO OTHER VARIABLES; ENGLISH SAMPLE

	1	2	3	4	5	6
Prejudice (F scale)	X	.29*	−.22	−.01	.20	.19
Preference (Bogardus)		X	−.40**	.17	.13	.13
Attitudes to French (incomplete sentences)			X	−.18	−.04	−.08
Attitudes to English (incomplete sentences)				X	.26*	.25
Favorableness of evaluations (English over French; all traits)					X	.77**
Favorableness of evaluations (English over French; desirable traits)						X

Note.—The N's vary around 55 for all entries except Variable 6 where they drop to around 50.

* Significant at .05 level.

** Significant at .01 level.

TABLE 4

RELATION OF ATTITUDINAL MEASURES TO OTHER VARIABLES; FRENCH SAMPLE

Variables	1	2	3	4	5	6
Prejudice (F scale)	X	.17	.03	−.07	−.05	−.11
Preference (Bogardus)		X	.38**	−.45**	.26*	.27
Attitudes to French (incomplete sentences)			X	−.43**	.25*	.45**
Attitudes to English (incomplete sentences)				X	−.21	−.20
Favorableness of Evaluations (French over English; all traits)					X	.72**
Favorableness of Evaluations (French over English; desirable traits)						X

Note.—The N's vary around 55 for all entries except Variable 6 where they drop to around 35.

Since the French *S*s in general showed more favorableness to English than French guises, Variables 5 and 6 mainly represent degrees of nonfavorableness to English guises.

* Significant at .05 level.

** Significant at .01 level.

as the comparative favorableness of evaluations to the spoken languages. For the English sample, the more bilingual subgroup had a reliably lower mean F score than did the less bilingual group, $t_{60} = 2.20$, a finding consistent with another study (Gardner & Lambert, 1958) on the relation of bilingualism to attitudes. The two bilingual groups did not differ reliably on any other variable, however. For the French sample, the only significant difference between the more and less bilingual subgroups was on the modified Bogardus scale: the more bilingual group showed reliably less own group preference than did those with little facility in English, $t_{52} = 2.43$.

DISCUSSION

Evaluational reactions. It is not unexpected that English speaking *S*s should show more favorableness to members of their linguistic group, but the finding that French *S*s also evaluate English guises more favorably is as unexpected as the finding that these *S*s judge French guises less favorably than do the English *S*s.

The traits given to *S*s for evaluation included several of those commonly considered necessary for social and economic success (e.g., looks, leadership, confidence, ambition), but we attempted to balance these with personality characteristics generally considered of greater value such as likability and kindness. It is possible that we were biased in the selection of traits and, through ignorance, presented a list which omitted characteristics which have value for French speaking Canadians. However, the French sample chose dependability, intelligence, and kindness as the most desirable traits for friends of those given them for consideration, and two of these were rated more favorably with the English guises, a fact which argues against the view that no opportunity was given the French sample to evaluate the speakers on culturally important traits.

In view of the greater probability of finding

English people in more powerful social and economic positions in the Montreal community, both samples of *S*s might more likely think of an English speaker as having higher status and then evaluate the English voice of the putative lawyer as having more ambition, intelligence, etc., than the French voice of the putative store employee. If the French *S*s had a greater tendency to do so, this might account in part for their evaluational reaction pattern. Data were available to test this notion in an indirect fashion by making use of *S*s' estimations of each speaker's likely occupation in his French and English guises. These occupational assignments were categorized as either professional or nonprofessional by two judges working independently. For the English sample, the ascribed status of English guises was markedly higher than for French guises, $\chi^2 = 38.14$ with 1 *df*. For the French sample, the same trend was noted but was much reduced in magnitude, $\chi^2 = 5.38$. The fact that the French *S*s ascribe higher status to a larger proportion of French guises (44%) than do English *S*s (33%) and also evaluate the French guises less favorably than do English *S*s argues against an interpretation in terms of differences in perceived status of the two groups.

The findings can also be interpreted as meaning that the French *S*s actually perceive members of the English speaking group as having various desirable personality traits. The characteristics ascribed to the English speakers are generally the same as those ascribed by the English *S*s to the same guises with the exception of the English-French conflict over which group is more kind. The French *S*s may well regard themselves as members of an inferior group, one that is nonetheless kind and religious. Several recent studies (Adelson, 1953; Sarnoff, 1951; Steckler, 1957) which document the manner in which minority groups sometimes adopt the stereotyped values of majority groups indicate that a parallel process may be taking place in the French Canadian community (cf. Stewart & Blackburn, 1956).

An examination of the common descriptions given by the French *S*s in the open-ended questionnaire suggests that they have ambivalent feelings about both French Canadians and English Canadians, an interpretation that is analogous to the attitude of American Jews toward both Jews and non-Jews as described by Lewin (1948, p. 145ff). The French *S*s describe their own group as progressing and having a potential (18 occurrences); feeling inferior or without confidence (16); having good qualities and being likable (16); being uninformed or uneducated (16); being intelligent or as intelligent as English Canadians (14); being individualistic or disunited (11); lacking ambition (11); and as being prejudiced or narrow-minded (10). They describe the English Canadians as being likable, good, or nice (27); thinking badly of French Canadians (19); feeling superior to French Canadians (13); being too Americanized (12); lacking knowledge of or appreciation for French Canadians (11); and as being less mannered or more rude (11).

The findings presented here take on somewhat more significance in view of two pilot studies using other samples of *S*s which show essentially the same results. In the first study, 17 English speaking college students listened to three of the four speakers used in the present study. In this case, English guises were evaluated more favorably on all five traits presented. In the second study, 22 male and 25 female French speaking members of Catholic religious orders were *S*s and three of the four speakers and 12 of the 14 scales used in the present study were administered. *S*s evaluated the English guises significantly more favorably on most traits except for kindness (*bonté*) on which French guises were more favorably evaluated. Speaker Tri, as in the present study, was not subject to this trend.

Attitude data. The general trend of correlations indicates that certain measures of attitude, particularly attitudes toward one's own group, are associated with the favorableness of evaluational reactions to the spoken languages, but this relation is clearly not a marked one, especially for the English *S*s. Furthermore, the prediction was not supported that more skill with the other group's language, which would permit more intimate interaction, would lead to smaller differences in favorableness of evaluational responses to the guises.

In view of the striking tendency for both English and French speaking *S*s to evaluate the English guises more favorably, a possible reason for the generally low correlations may be that generalized characteristics of French speaking and English speaking people are so widely accepted in the Montreal community that even those English *S*s with positive attitudes toward French may still perceive them as inferior on many traits. Likewise, French *S*s with negative attitudes toward English speaking people may still perceive them as superior in many respects. That is, both French and English *S*s regardless of their attitude may have come to believe that English speaking Canadians are taller and possibly more "intelligent" in the sense that they have more educational opportunities. To this extent, the attitude factor would be submerged by powerful community-wide stereotypes. This argument is limited, however, in that attitudes would certainly be expected to play the dominant role when the two guises are compared on such traits as good looks or character. To check this possible interpretation, correlations for the English sample were computed between the comparative favorableness of reactions to the guises and attitudes toward English Canadians for *each* trait examined separately. The correlations for character, intelligence, and dependability reach significance at about the .05 level but none of the others do, even those for good looks or kindness which would also be expected to reflect an attitude factor. We conclude, therefore, that the comparatively unfavorable perception of French speakers is essentially independent of the perceivers' attitudes toward French and English groups. The correlations for the French sample are not large or consistent enough to make French *S*s an exception to this generalization. Furthermore, experience with both groups, as inferred from the degree of skill in using both languages, also appears to be independent of the tendency to downgrade French speaking individuals. Further research is necessary to determine the influence of community-wide stereotypes and majority-minority feelings of members of the two language groups in Montreal.

Although the findings may be considered consistent and reliable, the generality of the results will become clear only after extensions to other samples of *S*s and a larger sample of speakers. It will be recalled that the four speakers used here were chosen to represent different linguistic communities: three speakers were educated in French Canadian universities but Leo spoke more naturally with a caricatured French Canadian accent. Only speaker Tri spoke French with a Parisian accent. A comparison of the evaluational reactions made to these speakers suggests that further study of accent differences of speakers from different linguistic communities may well extend the usefulness of the technique developed here. It can be seen in Table 1 that speaker Tri was not downgraded by the French *S*s when speaking French as were the other speakers; in fact he was considered significantly more self-confident in French than in English. Nor do the English *S*s treat him adversely in French, for they perceive him as having reliably more humor, entertainingness, and sociability in French than English. Clearly more than one speaker with a Parisian speech style is needed to test this notion, but the data suggest that the Montreal community may differentiate between Parisian and Canadian French in their attitudes. Note also in Table 1 that speaker Leo, the caricatured French Canadian, is particularly downgraded in his French guise by both English and French *S*s, if we interpret the size of *t* values as an index of differential favorableness of evaluation.

SUMMARY

Samples of French speaking and English speaking Montreal students were asked to evaluate the personality characteristics of 10 speakers, some speaking in French, others in English. They were not made aware that actually bilinguals were used as speakers so that the evaluational reactions to the two language guises could be matched for each speaker. The comparative favorableness of evaluations of the two guises was correlated with various measures of attitudes toward French and English Canadians for both samples of *S*s.

The results indicate that English *S*s evaluate the English guises more favorably on most traits. French *S*s not only evaluate the English guises more favorably than French guises, but their evaluations of French guises are reliably less favorable than those of English *S*s. This

finding is interpreted as evidence for a minority group reaction on the part of the French sample.

The correlations between comparative favorableness of English and French guises and measures of attitude toward own and other groups were generally low and insignificant, especially for the English sample. The essential independence of evaluational reactions to spoken languages and attitudes is interpreted as a reflection of the influence of community-wide stereotypes of English and French speaking Canadians.

REFERENCES

ADELSON, J. B. A study of minority group authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 477–485.

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

GARDNER, R. C., & LAMBERT, W. E. Social-motivational aspects of second-language acquisition. 1958. (Mimeo.)

LEWIN, K. *Resolving social conflicts.* New York: Harper, 1948.

LICKLIDER, J. C. R., & MILLER, G. A. The perception of speech. In S. S. Stevens (Ed.) *Handbook of experimental psychology.* New York: Wiley, 1951.

SARNOFF, I. Identification with the aggressor: Some personality correlates of anti-semitism among Jews. *J. Pers.*, 1951, **20**, 199–218.

STECKLER, G. A. Authoritarian ideology in Negro college students. *J. abnorm. soc. Psychol.*, 1957, **54**, 396–399.

STEWART, J., & BLACKBURN, J. Tensions between English-speaking and French-speaking Canadians. *Contributions à l'Etude des Sciences de l'Homme*, 1956, **3**, 145–167.

(Received October 20, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

ACHIEVEMENT MOTIVE AND TEST ANXIETY CONCEIVED AS MOTIVE TO APPROACH SUCCESS AND MOTIVE TO AVOID FAILURE

JOHN W. ATKINSON[1] AND GEORGE H. LITWIN[2]
University of Michigan

SINCE 1950 there have been an increasing number of studies using thematic apperceptive (or equivalent) measures of n Achievement (McClelland, Atkinson, Clark, & Lowell, 1953; Atkinson, 1958; French, 1958) and the Mandler-Sarason Test Anxiety Questionnaire (TAQ) (Mandler & Sarason, 1952; Sarason, Mandler & Craighill, 1952; Mandler & Cowen, 1958), which amply demonstrate that knowledge of motivational differences enhances prediction of achievement related performances. The reader who is familiar with this literature will have noted that studies of n Achievement show it to be a motive that generally enhances performance in achievement situations while studies of Test Anxiety show it to be a motive that normally produces decrements in achievement test performance. Interpretations of the results of studies that have employed only a measure of n Achievement have often included some reference to *S*s having low n Achievement scores as apparently more fearful of failure (Atkinson, 1953; Brown, 1953; Moulton, Raphelson, Kristofferson, & Atkinson, 1958). On the other hand, studies that have employed only the measure of Test Anxiety (Mandler & Sarason, 1952; Sarason & Mandler, 1952; Sarason, et al., 1952) have viewed the behavior of *S*s having low Test Anxiety scores as less conflicted and more task oriented. One might be tempted to think that these two sets of research findings overlap completely and that the measures employed merely tap opposite ends of a single motivational variable. This would seem a plausible inference in light of the correlation of −.43 between the two measures reported by Raphelson (1957) in a study that also showed each to be related in the expected direction to a physiological index of manifest anxiety in a stressful test situation.

The position taken here, however, is that the measure of n Achievement obtained from thematic apperception (McClelland, et al., 1953; Atkinson, 1958) or the French Test of Insight (French, 1955, 1958; French & Thomas, 1958), a very similar projective instrument, and the measure of Test Anxiety obtained from the self-knowledge scale developed by Mandler and Sarason are not measures of the same variable. Rather, it is assumed that these particular measures of n Achievement indicate the strength of a motive to approach success, while Mandler-Sarason Test Anxiety scores indicate the strength of a motive to avoid failure.

A theoretical model explaining how these two motives influence risk taking (i.e., selection) and level of performance in achievement test situations has been presented by Atkinson (1957, 1958) elsewhere. It provides a basis for testing assumptions about what is being measured by the two instruments. According to the theory, when a person's motive to achieve success is stronger than his motive to avoid failure, the resultant of the conflict is always positive; i.e., approach motivation, no matter what the level of difficulty of the task. An individual so motivated is most attracted to tasks of intermediate difficulty where the subjective probability of success is .50. Here the resultant positive motivation is strongest. However, if the motive to avoid failure (which is presumed to be a disposition to become anxious about failure under achievement stress) is stronger, then the resultant of the approach-avoidance conflict is avoidant motivation for all levels of difficulty. The maximum strength of avoidant motivation, i.e., the strongest anxiety about failure, occurs when tasks are of intermediate difficulty. Hence, according to the theory, a person so motivated finds all achievement tasks unattractive, particularly ones of intermediate difficulty.

---

[1] This research was supported in part by a grant from the Ford Foundation for Studies of Personality Dynamics for which the authors are very appreciative. A preliminary report of this study was given at the APA meeting in Washington in 1958.

[2] Now at Harvard University.

He performs them only when constrained by social pressures (otherwise he would "leave the field"),[3] and he performs them inefficiently; i.e., he suffers a performance decrement as a consequence of conflict engendered by competing avoidant tendencies, to the extent that his anxiety is aroused (Sarason, et al., 1952).

The present study tests several hypotheses derived from the theoretical model and simultaneously, the assumptions that (*a*) n Achievement, as measured by thematic apperception or a similar projective instrument, is a positive disposition to approach success and that (*b*) what has been called Test Anxiety is a disposition to avoid failure. The dependent variables are: goal setting (level of aspiration) when individuals are confronted with tasks which differ in difficulty (ring toss from distances ranging from 1 to 15 ft.); persistence at an achievement task when it is not clear that success has been attained (length of time spent working on a final examination); and efficiency of performance, or level of accomplishment (score on the final examination).

The hypotheses are as follows: Persons in whom the motive to achieve success is stronger than the motive to avoid failure (*a*) more often select tasks of intermediate difficulty, (*b*) work for a longer time on the final examination, and (*c*) get higher scores on the final examination than persons in whom the motive to avoid failure is the stronger motive. Hence, given the previously stated assumptions about the meaning of measures derived from the two tests, it is expected that n Achievement scores obtained from the French Test of Insight should be positively related to preference for intermediate risk, persistence, and efficiency of performance, and that scores from the TAQ should be negatively related to these same variables. Predictions should be enhanced when *S*s are simultaneously classified as high or low on both measures.

A secondary interest is to explore the the behavioral correlates of another measure called n Achievement which is obtained from the Edwards Personal Preference Schedule (1954). To date, there have been strong indications that direct (i.e., assent or preference) and indirect (i.e., imaginative or projective) methods of measuring achievement motivation do not yield comparable results. DeCharms, Morrison, Reitman & McClelland (1955) have found this to be true when a direct and indirect method of measuring n Achievement were compared in terms of predictions to behavior. Birney (see McClelland, 1958a) has found that measures of n Achievement obtained from thematic apperception and the Personal Preference Schedule (PPS) are uncorrelated ($r = -.002$, $N = 300$). This absence of relationship between the two measures is also reported by Marlowe (1959) who found, in addition, that the thematic apperceptive measure of n Achievement was positively related to peer group ratings of achievement related behavior, while the measure obtained from the PPS was not.

In summary, the purpose of this investigation is to examine the construct validity (see Cronbach & Meehl, 1955) of three contemporary measures of achievement related motives. Predictions concerning the behavioral correlates of the measures employed are generated by a theory that conceives the achievement motive and the motive to avoid failure as independent, latent, directional dispositions and states the conditions *under* which and the degree *to* which they are aroused and manifested in overt action.

METHOD

Subjects

The *S*s were male students enrolled in a sophomore-junior level psychology course at the University of Michigan in the fall of 1957. The data were collected in the laboratory sections of the course. The instructors served as *E*s.[4] There were 49 *S*s on whom both the measures of n Achievement and Test Anxiety were available. Not all of these *S*s, however, appeared for all of the later data collection sessions. In all comparisons, the largest possible number of *S*s is used, even though this results in slight fluctuations in the size of samples among the various comparisons.

Measure of n Achievement and Test Anxiety

Early in the semester *S*s were administered, under neutral classroom conditions, a Test of Insight developed by Elizabeth French (1958). The test consists of 10 short verbal statements about people which *S*s are asked to analyze. Both Forms I and II of the Test of Insight were used, half of the *S*s getting each form. The interpretations of behavior, which are very similar to TAT stories, were scored for n Achievement (according to the method of content analysis described by McClelland, et al. 1953; Atkinson, 1958) by a scorer whose reliability had been established above .90 on training materials.[5] The distribution of n Achievement scores for each form was broken at the median. *S*s above the median were classified as high in n Achievement and those below as low in n Achievement.

---

[3] "Social pressures" reduce to other kinds of motivation, extrinsic to achievement, for performing the tasks (e.g., desire for social approval), which are not normally measured in experiments of this sort.

[4] We wish to acknowledge the assistance of Joseph Veroff and Lawrence Littig.

Test Anxiety scores were obtained from the Mandler-Sarason TAQ (see Mandler & Cowen, 1958). The TAQ were distributed in class and *S*s were asked to complete them outside of class and to return them the following week. The questionnaires were scored by dividing each scale into intervals, assigning a score of 1–4 for each scale, and summing the scale scores to obtain the Test Anxiety score. This scoring method correlates very highly with that originally described by Mandler and Sarason ($r = .94$, $N = 50$), and differs from that recently described by Mandler and Cowen only in that it uses 4 intervals instead of 10. The distribution of Test Anxiety scores was dichotomized at the median. The PPS developed by Edwards (1954) was administered to *S*s in class. The distribution of PPS n Achievement scores was also dichotomized at the median.

Goal Setting in Modified Ring Toss Game

Several weeks after the individual difference measures had been obtained, *S*s played a ring toss game. Four games were set up in the room when they entered, one on each side, one in the front, and one in the back of the room. Each game consisted of a wooden peg 2 in. in diameter and 12½ in. high mounted on a round wooden base, a ring 10 in. in diameter, and 15 lines marked on the floor at one ft. intervals. A number was marked on the floor next to each line. The closest line was one ft. from the target and numbered 1; the farthest line was 15 ft. from the target and numbered 15.

The *S*s were told: "Today you are going to play a ring toss game. You will have an opportunity to take 10 shots at the target from any line you wish. You may move after each shot or shoot from the same line. Someone will record your shots and get your code number when you finish. We want to see how good you are at this." The instructor then kept track of the shots and hits of the first few *S*s, and these *S*s then kept score for the rest of the class. Two of the ring toss games were used by men and two by women. There was a good deal of informal banter during the session. No attempt was made to interfere with or to control the lifelike situation. Only the results for men are considered here since there remain unresolved questions concerning the validity of measures of n Achievement and Test Anxiety when applied to women.

Data from Final Examination

The final examination in the course from which the *S*s were drawn was a multiple choice and short answer test held several months later. It was scheduled to last from 9:00 to 12:00 a.m. Most of the students did not take the full three hours. As each student left the examination room after turning in his exam, and when he was out of sight of students still working on the exam, he was handed a card with the time written on it and asked to write his code number on it. The code number made it possible to relate the amount of time spent working on the exam to the individual difference measures, to goal setting behavior in the ring toss game, and to the final examination scores. The data were transcribed into a score for each *S* which represented the number of minutes spent working on the final examination.[6]

RESULTS

Descriptive analysis of goal-setting data

The *S*s were classified simultaneously as high or low n Achievement and Test Anxiety. It follows from our assumptions about the two measures that the group classified high n Achievement-low Test Anxiety should be more strongly motivated to approach success than to avoid failure. The low n Achievement-high Test Anxiety group, on the other hand, should be more strongly motivated to avoid failure than any other group. The other two groups, the high n Achievement-high Test Anxiety group and the low n Achievement-low Test Anxiety group should be more conflicted or intermediate in resultant motivation.

Figure 1 presents a smoothed curve for each of the four groups describing the percentage of shots taken from each distance. The curves are smoothed by the method of running averages. From Fig. 1 it can be seen that all four groups show some preference for the distances around 9-11 feet. However, *S*s high in n Achievement and low in Test Anxiety prefer this region most strongly, while the low n Achievement–high Text Anxiety group shows the weakest preference for it. The other two groups are between these two as predicted. At both the closest and farthest distances, the low n Achievement Test Anxiety group takes more shots than the high n Achievement-low Text Anxiety group. The curves indicate a relative difference

---

[5] We wish to acknowledge the assistance of Lois Hendrickson who scored the protocols. Because many of the interpretations had no achievement imagery, the Unrelated Imagery category (UI) was scored 0 instead of −1.

[6] Because of overflow in the main examination room, 20 to 30 students took the examination in a different room. A constant of 5 min. was subtracted from the score of these students since they began the examination about 5 min. later.

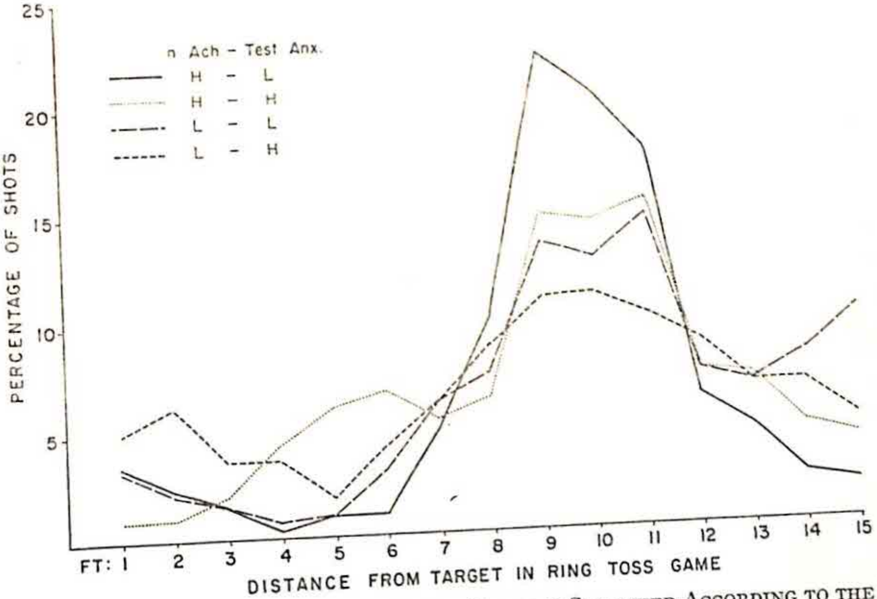


Fig. 1. Percentage of Shots Taken from Each Line. Graph is Smoothed According to the Method of Running Averages, for *S*s Classified as High or Low Simultaneously in n Achievement and Test Anxiety, H-L ($N = 13$), H-H ($N = 10$), L-L ($N = 9$), L-H ($N = 13$)

between the motivation groups in preference for intermediate difficulty.

In Table 1, various ways of analyzing the data into easy, intermediate, and difficult regions are explored. The first method (I) is to divide the range of distances into equal geographic thirds. This is the method employed by Atkinson, et al. (1959) in an earlier study. The second method (II) is to use the obtained distribution of shots to determine the easy, intermediate, and difficult regions. Use of the obtained distribution of shots to define degrees of difficulty was suggested by McClelland (1958b) in an investigation of risk taking in children. The obtained distribution was first broken into approximate thirds (II-A) and then into the approximate interquartile range and extreme quarters (II-B) to show the effect of widening the region of intermediate difficulty. The third method (III) takes the middle third of distance around the obtained median of 9.8 feet.

Examination of Table 1 shows that all of the methods yield comparable results. The high n Achievement-low Test Anxiety group consistently takes the greatest percentage of shots from the intermediate region and the low n Achievement-high Test Anxiety group consistently takes the least number of shots from the intermediate region The difference between the percentage of shots taken by these two groups from the intermediate distances ranges from 27% to 34%. The high n Achievement-high Test Anxiety group and the low n Achievement-low Test Anxiety group fall between the two extreme groups as expected.

Effects of n Achievement and Test Anxiety

It is desirable to express the goal setting tendency of each person as a single score in statistical tests. An Average Deviation score, which was found by Litwin (1958) to be the most consistent measure across tasks of the tendency of persons to prefer intermediate risk, was used. An Average Deviation score was computed for each shot as follows: Distance of shot-Mdn. distance of all shots/Average Deviation of all shots. The Average Deviation score for each *S* was the mean of his scores (irrespective of sign) on the 10 shots. A low Average Deviation score means that shots were taken from intermediate distances; a large score means that shots were taken from very close or very far back in relation to the median of all shots.

In Table 2 the performance of *S*s high and low in n Achievement is summarized for the three dependent variables: ring toss Average

TABLE 1

PERCENTAGE OF TEN RING-TOSS SHOTS TAKEN BY *S*s SIMULTANEOUSLY CLASSIFIED HIGH AND LOW IN n ACHIEVEMENT AND IN TEST ANXIETY USING ALTERNATIVE CRITERIA FOR DEFINITION OF DEGREE OF DIFFICULTY (OR RISK)

Basis for Definition of Three Degrees of Difficulty (or Risk)	Distance in Feet	Motivation (n Achievement Test Anxiety)			
		High-Low $N = 13$	High-High $N = 10$	Low-Low $N = 9$	Low-High $N = 13$
I. Using geographic distance: equal thirds	1–5	8%	15%	9%	21%
	6–10	70	55	49	43
	11–15	22	30	42	36
II. Using obtained distribution of shots: A. approximate thirds	1–8	22	31	28	39
	9–10	55	39	30	25
	11–15	22	30	42	36
B. approximate interquartile range vs. extreme quarters	1–7	11	26	18	32
	8–11	73	48	48	41
	12–15	16	26	34	28
III. Using both obtained distribution of shots and geographic distance: middle third of distance about the obtained median shot (9.8 ft.)	1–7	11	26	18	32
	8–12	82	60	58	48
	13–15	7	14	24	20

TABLE 2

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *S*s CLASSIFIED HIGH OR LOW IN n ACHIEVEMENT AND PERCENTAGE OF *S*s ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	n Achievement		High vs. Low[a]
		High	Low	
Risk taking: average deviation score in ring toss game	N	23	22	
	Mdn.	.63	.91	$U = 177.5$
	% below combined group Mdn.	61	36	$p = .04$
Persistence: minutes spent working on final exam.	N	25	19	
	Mdn.	164.9	154.0	$U = 156.5$
	% above combined group Mdn.	60	32	$p = .03$
Efficiency: final exam score	N	25	19	
	Mdn.	101	85	$U = 152.5$
	% above combined group Mdn.	64	32	$p = .02$

Note.—The slight fluctuation in N is explained in the Method section.

[a] Mann-Whitney U Test. All tests of significance are one-tailed tests.

Deviation score; minutes spent working on the final exam; and the score on the final exam. From Table 2 it can be seen that the high n Achievement group tends to have lower Average Deviation scores, i.e., to show a greater preference for intermediate risk, to spend more time working on the final exam, and to have higher final examination scores than the low n Achievement group. The differences between the two groups were tested by the Mann-Whitney U test. Since the direction of differences was predicted in each case, one-tailed tests were made. The differences between the high and low n Achievement groups are all statistically significant. Table 2 also shows the percentage of *S*s falling above or below the combined group median in each case; 60-64% of the high n Achievement group fall in the predicted cell, and 64-68% of the low n Achievement group fall in the predicted cell.

Table 3 presents the same information for the high and low Test Anxiety groups. Again, all the results are as predicted. The high Test Anxiety group tends to have higher Average Deviation scores, i.e., to show greater avoidance of intermediate risk, to leave the final exam sooner, and to have lower final examination scores than the low Test Anxiety group. Mann-Whitney U tests were computed on the differences between the motivation groups;

TABLE 3

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *Ss* CLASSIFED HIGH OR LOW IN TEST ANXIETY AND PERCENTAGE OF *Ss* ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	Test Anxiety		High vs. Low[a]
		High	Low	
Risk taking: average deviation score in ring toss game	*N* Mdn. % below combined group Mdn.	23 1.00 35	22 .63 64	$U = 177$ $p = .04$
Persistence: minutes spent working on final exam	*N* Mdn. % above combined group Mdn.	22 154.8 32	22 165.2 64	$U = 176$ $p = .06$
Efficiency: final exam score	*N* Mdn. % above combined group Mdn.	22 91.0 41	22 101.5 59	$U = 177$ $p = .06$

Note.—The slight fluctuation in *N* is explained in the Method section.
[a] Mann-Whitney *U* Test. All tests of significance are one-tailed tests.

TABLE 4

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *Ss* SIMULTANEOUSLY CLASSIFIED HIGH OR LOW IN n ACHIEVEMENT AND TEST ANXIETY AND PERCENTAGE OF *Ss* ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	n Achievement Test Anxiety				High-Low vs. Low-High[a]
		High-Low	High-High	Low-Low	Low-High	
Risk taking: average deviation score in ring toss game	*N* Mdn. % below combined group Mdn.	13 .48 77	10 1.03 40	9 .88 44	13 .93 31	$U = 42.5$ $p < .025$
Persistence: minutes spent working on final exam	*N* Mdn. % above combined group Mdn.	15 165.3 73	10 157.5 40	7 148.3 43	12 154.5 25	$U = 40.5$ $p < .01$
Efficiency: final exam score	*N* Mdn. % above combined group Mdn.	15 102.3 67	10 96.5 60	7 91 43	12 82.5 25	$U = 47.5$ $p < .025$

Note.—The slight fluctuation in *N* is explained in the Method section.
[a] Mann-Whitney *U* Test. All tests of significance are one-tailed tests.

one-tailed tests were made. The groups differ significantly in Average Deviation scores ($p = .04$). For both the number of minutes spent working on the final exam and final examination scores the probability of the difference in the predicted direction is .06. It can be seen that 59-64% of the low Test Anxiety group, and 59-68% of the high Test Anxiety group fall in the predicted cells in relation to the combined group median.[7]

In Table 4, *Ss* are again simultaneously classified on n Achievement and Test Anxiety. The high n Achievement-low Test Anxiety group is seen to have the lowest Average Deviation score (i.e., the greatest preference for intermediate risk), to spend the most time working on the final exam, and to have the highest final examination scores. The low n Achievement-high Test Anxiety group tends to have the highest Average Deviation scores, (i.e., the weakest preference for intermediate risk), to leave the final exam the earliest, and to have the lowest final examination scores. The high n Achievement-high Test Anxiety group and the low n Achievement-low Test Anxiety group generally fall between the two extreme groups as predicted. Mann-Whitney

---

[7] When only the extreme quartiles are compared, as is common in research using the Test Anxiety Questionnaire, the percentage of groups ($N = 11$) falling into predicted cells is increased somewhat to 63–72%. The median split is used here both because the *N* is relatively small to begin with and also to call attention to the relative sensitivities of the various measures of achievement related motivation when applied to the same dependent variables.

TABLE 5

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *S*s CLASSIFIED HIGH OR LOW IN PPS n ACHIEVEMENT AND PERCENTAGE OF *S*s ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	PPS n Achievement		High vs. Low[a]
		High	Low	
Risk taking: average deviation score in ring toss game	*N*	22	21	
	Mdn.	.97	.50	$U = 137.5$
	% below combined group Mdn.	36	62	$p = .02$
Persistence: minutes spent working on final exam	*N*	24	20	
	Mdn.	157.5	163.0	$U = 223.5$
	% above combined group Mdn.	42	55	$p = .70$
Efficiency: final exam score	*N*	24	20	
	Mdn.	100.0	89.5	$U = 181.5$
	% above combined group Mdn.	58	40	$p = .17$

Note.—The slight fluctuation in *N* is described in the Method section.
[a] Mann-Whitney *U* Test. All tests of significance are two-tailed tests.

U tests of the differences between the high n Achievement-low Test Anxiety group and the low n Achievement-high Test Anxiety group employing one-tailed alternative hypotheses are all statistically significant. When the percentage of *S*s in each of the motivation groups falling above or below the combined group median is computed, 67-77% of the high n Achievement-low Test Anxiety group are correctly placed, and 69-75% of the low n Achievement-high Test Anxiety group are correctly placed. Simultaneous classification on both approach and avoidance motives clearly improves the prediction.

The examinations were graded A+ (excellent) to E (failure) before this analysis was even begun. It is of interest to note that the median final examination score for the high n Achievement-low Test Anxiety group (102.3) was a B− grade on the final exam, and that 67% of this group received B− or better on the exam while only 7% received a grade of D or E. On the other hand, the median final examination score for the low n Achievement-high Test Anxiety group (82.5) was a D grade; 58% of this group received a D or an E on the final exam and only 25% received a grade of B− or better.[8]

The Measure Called n Achievement on the PPS

Table 5 presents the results for *S*s classified as high or low on PPS n Achievement. The only statistically significant result is for the ring toss Average Deviation scores, and here it is directly opposite to what would be expected if the test measured the strength of motive to achieve success as conceived in the risk taking theory. The high PPS n Achievement group have significantly higher Average Deviation scores, (i.e., greater avoidance of intermediate risk) than the low PPS n Achievement group ($p = .02$). Two-tailed alternative hypotheses were used here since there is little evidence in previous research to support any assumptions about what this test measures. For neither time spent working on the final examination nor final examination scores are the differences between the two PPS n Achievement groups statistically significant. The trends are inconsistent. If anything, *S*s having high PPS n Achievement scores do not work as long on the exam but get better grades than those having low scores.

---

[8] These results are viewed as consistent with the idea that highly anxious individuals suffer a performance decrement on achievement tests (Sarason, et al., 1952). It may be argued that unmeasured differences in ability and not differential anxiety are responsible for the result. Perhaps *S*s score high on Text Anxiety because they have performed badly in the past (also as a consequence of low ability). The argument is plausible, but so also is the alternative. Accordingly, finding that *S*s who score high on Text Anxiety also score low on some independent test of academic ability would be expected since ability is always measured under the very kind of stressful test conditions that produce performance decrements in anxious *S*s. Experimental evidence (e.g., Sarason, et al., 1952) which shows that anxious *S*s perform as well or better than nonanxious *S*s under nontest conditions but more poorly under achievement test conditions makes the motivational interpretation plausible pending the discovery by someone of a method of isolating the motivational and aptitudinal influences on the usual so-called test of ability.

Since the PPS n Achievement measure behaves as if it were a measure of the motive to avoid failure (at least in the risk taking game), the PPS n Achievement measure was substituted for the Test Anxiety measure in a simultaneous classification with n Achievement (French Test of Insight). The high n Achievement-low PPS n Achievement group ($N = 11$) had 82% of its *S*s below the combined group median for Average Deviation score, the low n Achievement-high PPS n Achievement group ($N = 10$) had 30% of its *S*s below the combined group median. The difference between these two extreme groups was very significant ($U = 16, p = .01$ in this direction), larger than the comparable difference shown in Table 4 when Test Anxiety is used to assess the tendency to avoid failure!

When *S*s were classified on all three of the measures, n Achievement, Test Anxiety, and PPS n Achievement, striking differences were noted. The high n Achievement-low Test Anxiety-low PPS n Achievement group ($N = 7$) had 100% of its *S*s below the combined group median for Average Deviation score while the low n Achievement-high Test Anxiety-high PPS n Achievement group ($N = 8$) had only 25% of its members below the combined group median. The difference between these two groups was highly significant ($p = .001$).

Intercorrelation of Variables

Spearman rank order correlations were used to ascertain the degree of intercorrelation among the independent and dependent variables. The three measures of achievement related motives were not significantly related. The correlations between the several measures were as follows: Test Anxiety and PPS n Achievement, .11 ($N = 47$); PPS n Achievement and n Achievement, $-.05$ ($N = 47$); Test Anxiety and n Achievement, $-.15$ ($N = 47$).

The three dependent behavioral variables, however, tended to be positively related to one another. The index of risk taking preference, Average Deviation score (low scores given high ranks), and the measure of persistence, time spent working on the final exam, were correlated .34 ($N = 40, p < .05$); number of min. spent working on the final exam and final examination score were correlated .27 ($N = 40, p < .10$); the Average Deviation score (low scores given high ranks) and final examination scores were correlated .19 ($N = 40, p < .25$). A more substantial relationship is evident between the two behavioral indices of motivation, risk taking and persistence, that are not strongly influenced by academic ability, than between either of these measures and exam performance which is strongly influenced by ability.

DISCUSSION

Construct Validity of Three Tests

These results yield support for the assumptions stated earlier regarding the nature of the variables being measured by the French Test of Insight and the Mandler-Sarason TAQ. Since Litwin (1958) reports directly comparable risk taking results when using thematic apperception to assess achievement motivation, it can be concluded that n Achievement scores obtained from thematic apperception or the Test of Insight indicate the strength of a motive to approach success. Test Anxiety scores yield results consistent with the assumption that the variable involved is a disposition to avoid failure. This assumption is not at variance with Mandler and Sarason's (1952; Sarason, et al., 1952) explanation of the performance decrement shown by anxious persons under achievement stress: the anxious person's performance suffers when task-irrelevant (avoidant) responses interfere with task-relevent (approach) responses.

The results concerning efficiency of performance, or level of accomplishment, under achievement-oriented conditions replicate earlier findings concerning n Achievement (Atkinson & Reitman, 1956; French, 1955) and Test Anxiety (Mandler & Sarason, 1952; Sarason, et. al., 1952). The positive relationship between n Achievement and persistence at a task replicates an earlier finding by French and Thomas (1958) and is further evidence of the tendency also noted in recall of interrupted tasks (Atkinson, 1953). To our knowledge, this is the first study to relate Test Anxiety to a measure of persistence. The relationship of Test Anxiety to risk taking or goal setting behavior was first noted by Vitz (1957), and Litwin (1958) has recently replicated the present risk taking findings using several different tasks.

The results concerning the measure of n Achievement obtained from the Edwards PPS are nebulous. This is the third occasion on which no relationship has been found between this direct preference measure and an indirect or projective measure of n Achievement. The observed tendency of persons having high PPS n Achievement scores to avoid intermediate risk, as if they were in fact more strongly motivated to avoid failure than to achieve, dramatically reinforces the statement that *different methods of assessing human motives do not yield comparable results* (see McClelland, 1958a). The inconsistency and lack of statistical significance of the other results pertaining to the PPS measure of n Achievement make it very difficult to understand just what is being measured by this test. One suggestive lead is provided by some unpublished results obtained by Charles Seashore.[9] He had *S*s judge the social desirability of all the statements on the PPS on a seven point scale about three months after taking the test and found a correlation of .47 ($N = 50$) between PPS n Achievement score and judged social desirability of the achievement statements. One might conclude from this that although the paired items are equated for *average* social desirability, individual differences in scores obtained from this test can be considered measures of individual differences in what is deemed socially desirable. This conclusion is supported by a recent study by Corah, Feldman, Cohen, Gruen, Meadow, and Ringwall (1958) in which judged social desirability of items within pairs on the PPS correlated highly with the item selected. Arguing in a similar vein, McClelland (1958a) has suggested designating measures obtained by direct preference or assent methods as "values", e.g., *v* Achievement (see also deCharms, et al., 1955), in order to avoid the inevitable confusion that arises in trying to reconcile research findings involving variables assessed by different methods but gratuitously called by the same name.

In terms of the theoretical conception of risk taking and achievement related behavior presented in detail elsewhere (Atkinson: 1957, 1958) and the assumptions regarding the variables measured here, the present findings provide considerable evidence of construct validity for the n Achievement scores obtained from the French Test of Insight and for Test Anxiety scores interpreted as measures of the strength of a disposition to be anxious about, and hence motivated to avoid, failure. However, until national norms are available for each of these measures, the crude designations as "high" and "low" of groups who stand above or below the median score in a college population are at best crude estimates of relative strength.

The relationship of n Achievement and Test Anxiety

The zero order correlation between n Achievement and Test Anxiety reported here is at variance with Raphelson's (1957) earlier report of a negative correlation. Yet the results may not, in fact, be inconsistent. In the present study the projective measure of n Achievement was presented under *neutral* classroom conditions, and the results show that persons having high n Achievement scores are as likely as not also to have high Test Anxiety scores. In Raphelson's study, the n Achievement scores were obtained from thematic apperceptive stories written by *S*s under *achievement orientation* during a rest period in the middle of a rather stressful test situation. The results of Martire (1956) and Scott (1956) clearly point to the possibility that persons who are strong in *both* achievement motive and motive to avoid failure may inhibit or distort the expression of achievement related imagery when anxiety about failure has been actively aroused in them by threatening situational cues. If so, it would be expected that persons classified high n Achievement (on the basis of *neutral* thematic apperceptive or Test of Insight scores) and also high in Test Anxiety should express less achievement related imagery when anxiety about failure is aroused. This state of affairs would induce the negative correlation between the measures reported by Raphelson.

---

[9] Personal communication, Sept., 1958.

Given only high (above the median) n Achievement scores, one can assume that the achievement motive is relatively strong in relation to an unmeasured motive to avoid failure; given low n Achievement scores, one can assume that the motive to achieve success is relatively weak in relation to an unmeasured motive to avoid failure. When the motive to avoid failure is also measured, using the Test Anxiety Questionnaire, and *S*s having high n Achievement scores but also having high Test Anxiety scores are segregated, we then have more reason to be confident that our inference of relatively stronger motive to achieve success in the remaining high n Achievemet-low Test Anxiety group is correct, as the present results indicate. Similarly, segregating *S*s who have low scores on both n Achievement and Test Anxiety leaves a group low in n Achievement and high in Test Anxiety in whom we have more reason to be confident that the motive to avoid failure is relatively stronger. Yet in a group of college students it is unlikely that these are all persons in whom the absolute strength of motive to avoid failure is greater than the absolute strength of achievement motive. We find that this group does not actually *avoid* intermediate risk, as would be expected if the avoidance motive were stronger than the approach motive. Instead we find only a relatively weak preference for intermediate risk, in comparison with other groups. Presumably, most persons in whom the avoidant motive is actually stronger in an absolute sense are eliminated long before they reach college. The findings of Miller (1951), Eysenck (1955), and Eysenck and Himmelweit (1946) concerning extremely high levels of aspiration set by persons diagnosed as neurasthenic and extremely low levels of aspiration among persons diagnosed as hysteric represent the pathological extremes in avoidant motivation which may appear very rarely in college groups. The present research used all available *S*s and made its cutting point in the distributions at the median. Perhaps the choice of more extreme groups of *S*s, as has been common in research with the Mandler-Sarason test, would produce such avoidant groups even in a college population. The small number of *S*s in the present investigation precluded this possibility.

The Measure of Subjective Probability of Success

An issue of fundamental importance that is treated rather inadequately here is the measurement of subjective probability of success. Strong assumptions are made concerning how the chances of success must appear to *S*s as they approach the ring toss game. It is implicitly assumed here, as in an immediately preceding study (Atkinson, et al., 1959) which made the point explicit, that the probability of success must have appeared closer to 1.00 than to .50 at 1 ft. and closer to 0 than to .50 at 15 feet. The zone of intermediate risk or intermediate difficulty, i.e., where probability of success approximates .50, must therefore fall somewhere in between.

The actual probabilities of success obtained from the shots taken by *S*s in this experiment, while not altogether relevant since it is probability of success as it appears to *S*s at the beginning of the game that is required to test hypotheses from the risk taking theory, do nevertheless provide some support for the assumptions made. At 1 ft., the actual probability of success was 1.00. At 15 ft., it was 0. And at 7 ft., the actual probability of success was .52. However, the average observed probability of success was only .23 between 8 and 12 ft. where most shots were taken by the high n Achievement-low Test Anxiety group.

Some data collected by Litwin (1958) in a follow-up study illuminate the problem somewhat. Litwin found that when *S*s were asked to estimate their chances of hitting from each line in an 18 ft. game before performing, the average reported probability of success was nearest to .50 at 11-12 feet. This was two-thirds of the distance from the easiest to the most difficult line. In the present experiment, the modal point of shots by the high n Achievement-low Test Anxiety group was 9-10 ft., two thirds of the distance from 1 ft. to 15 feet. The *S*s in Litwin's study approached the game with overly optimistic expectancies of success. In all likelihood the same thing occurred in the present experiment, though no evidence can be presented to support the claim.

SUMMARY

It is assumed that measures of n Achievement obtained from the French Test of Insight (and thematic apperception) indicate the strength of a motive to achieve success and Mandler-Sarason Test Anxiety scores indicate the strength of a motive to avoid failure. Given these assumptions and a theory of motivational determinants of risk taking behavior, the following hypotheses are investigated: persons in whom the motive to achieve success is stronger than the motive to avoid failure (*a*) should prefer tasks of intermediate difficulty, (*b*) should show greater persistence in working at an achievement related task, and (*c*) should show more efficiency, or a higher level of accomplishment, than persons in whom the motive to avoid failure is stronger than the motive to achieve success.

Results on 49 college men regarding distance of shots in a ring toss game, time spent working on a final examination, and score on the final examination support the hypotheses and hence the construct validity of the two measures of motive strength.

Relationships between these behavioral variables and the variable called n Achievement on the PPS do not support the inference that strength of motive to achieve success is being measured by this test.

The three measures of achievement related motives are not correlated. These results highlight the importance of discovering why different methods of measuring apparently the same human motive do not yield comparable results.

REFERENCES

ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.* 1953, **46**, 381–390.

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

ATKINSON, J. W., BASTIAN, J. R., EARL, R. W., & LITWIN, G. H. The achievement motive, goal-setting, and probability preferences. *J. abnorm. soc. Psychol.*, 1960, **60**, 27–36.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.

CORAH, N. L., FELDMAN, M. J., COHEN, IRA S., GRUEN, W., MEADOW, A., & RINGWALL, E. A. Social desirability as a variable in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1958, **22**, 70–72.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

DECHARMS, R. C., MORRISON, H. W., REITMAN, W. R., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955.

EDWARDS, A. L. *Edwards Personal Preference Schedule.* Manual. New York: Psychological Corp., 1954.

EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.

EYSENCK, H. J., & HIMMELWEIT, H. T. An experimental study of the reactions of neurotics to experiences of success and failure. *J. gen. Psychol.*, 1946, **35**, 59–75.

FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.

FRENCH, ELIZABETH G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

FRENCH, ELIZABETH G., & THOMAS, F. H. The relation of achievement motivation to problem-solving effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56**, 45–48.

LITWIN, G. H. Motives and expectancy as determinants of preference for degrees of risk. Unpublished honors thesis, Univer. of Michigan, 1958.

MCCLELLAND, D. C. Methods of measuring human motivation. In J. W. Atkinson, (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. (a)

MCCLELLAND, D. C. Risk-taking in children with high and low need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. (b)

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MANDLER, G., & COWEN, JUDITH E. Test anxiety questionnaires. *J. consult. Psychol.*, 1958, **22**, 228–229.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MARLOWE, D. Relationships among direct and indirect measures of the achievement motive and overt behavior. *J. consult. Psychol.*, 1959, **23**, 329–332.

MARTIRE, J. G. Relationships between the self concept and differences in strength and generality of achievement motive. *J. Pers.*, 1956, **24**, 364–375.

MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, **46**, 378–387.

MOULTON, R. W., RAPHELSON, A. C., KRISTOFFERSON, A. B., & ATKINSON, J. W. The achievement motive and perceptual sensitivity under two conditions of motive arousal. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

RALPHELSON, A. C. The relationship between imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54,** 13–18.

SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47,** 810–817.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47,** 561–565.

SCOTT, W. A. The avoidance of threatening material in imaginative behavior. *J. abnorm. soc. Psychol.*, 1956, **52,** 338–346.

VITZ, P. The relation of aspiration to need achievement, fear of failure, incentives, and expectancies. Unpublished honors thesis, Univer. of Michigan, 1957.

(Received October 21, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SOME EFFECTS OF IMMEDIATE VERSUS RANDOMLY DELAYED SHOCK ON AN INSTRUMENTAL RESPONSE AND COGNITIVE PROCESSES

M. R. D'AMATO AND W. E. GUMENIK

New York University

It is a common assumption that, other things equal, an unpredictable aversive stimulus has greater aversive properties than a predictable one. This general assumption often takes the more specific form that an unpredictable shock is more unpleasant, produces more anxiety, etc. than a predictable shock of equal intensity. In spite of (or perhaps because of) the rather widespread acceptance of this view, relevant data on the human level are almost nonexistent (Sidman & Boren, 1957). Confirming evidence would have an importance aside from demonstrating the obvious. It might suggest a "baseline" situation in which variables relating to individual susceptibility to both types of aversive stimulation could be studied in relation to personality variables. The experimental situation described below offers promise in this connection.

METHOD

Subjects. The *S*s were 21 male college students enrolled in the introductory psychology class.

Apparatus. The apparatus consisted of a panel on which were mounted three 7-watt lights and two response levers. A centered green light served as the "starting" stimulus, indicating the beginning of a trial. There was a right and a left response lever, each 6 in. off the midline, and a right and a left white "reinforcing" light 10 in. above the appropriate lever.

Procedure. The *S* was seated in front of the panel and orally given standard instructions, the relevant portions of which follow:

> This is a study of guessing behavior. Part of the experiment involves shock, so I'll have to put these electrodes on.
>
> When this green light goes on, you are to guess which of these two white lights will go on. Indicate your guess by pressing the lever below the light which you think will go on. This is like a gambling situation in the sense that which light goes on on any trial is determined by chance, so all you can do is guess. . . .
>
> You will be shocked during the second part of the experiment, but not during the first part. Once the shocks start, one will occur on every trial for the remainder of the experiment. Any questions?

The shock, delivered through an Applegate constant current stimulator, was then administered and adjusted to a setting which *S* sound "unpleasant but not painful." The *S* was finally requested to advise *E* if the shock became "too strong or too weak" during the course of the experiment, so that it could be readjusted. Actually, unknown to *S*, the shock was raised slightly every fifth trial of the shock portion of the experiment, a maximum of 5 milliamps. being reached in about 25% of the cases.

There were 90 trials in all. Trials 1—60 did not involve shock. The sequence of events on these trials was: green (starting) light on; as soon as *S* responded, the green light extinguished and either the right or left white light went on for 1 sec., indicating the correct response. Thirty seconds later another trial started with the onset of the green light.

Trials 61–90 involved the addition of shock on every trial, but were otherwise the same as the previous trials and were contiguous with them. One lever ("I" lever) had been preassigned to immediate shock (IS): every time *S* pressed this lever he received a shock (on the left hand) immediately after the reinforcing light extinguished, irrespective of whether he was correct or incorrect on that trial. Pressing the other lever ("D" lever) always resulted in a randomly delayed shock (RDS) of the same intensity. The delays were 0, 10, and 20 sec., occurring with relative frequencies of .20, .40, and .40, respectively. The position of the I and D levers was randomly assigned to each *S*, with the restriction that for half of the *S*s the I lever was on the right.

Throughout the experiment, the right lever was correct 50% of the time for all *S*s. Thus, regardless of *S*'s strategy, his expected percentage of correct responses was 50%. This is an important aspect of the design, since it provides for a baseline from which the effects of the independent variable may be measured. It was assumed that, by the end of the first part of the experiment (Trials 1–60), most *S*s would be responding to each lever about equally often, conforming to the reinforcement pattern prevailing. Hence, if the RDS should prove more aversive than the IS, the *S*s should press the D lever less frequently than the I lever during Trials 61–90. Of course, since the "control" nonshock Trials 1–60 are available, one could conceivably use most any distribution of reinforcement, e.g., 60:40. However, to do so could densensitize the whole procedure, since the *S*s would no doubt distribute their responses during Trials 1–60 about 60:40, and the preference for the one lever so instituted might interact with the shock variable on Trials 61–90.

At the termination of the experiment *S* was read the following questions, one by one, and his answers secured:

1. (*a*) Before shock was introduced, on what percentage of the trials did the left light go on?
   (*b*) After shock was introduced?
2. (*a*) Before shock was introduced, on what per-

centage of the trials did you press the left lever?

(*b*) After shock was introduced?

3. You probably noticed that responses to one of the levers were followed by a quick shock, while responses to the other lever resulted in a shock which was somewhat delayed for a while. If you did notice, which lever was associated with the delayed shock?

4. (*a*) Were both types of shock equally unpleasant?
   (*b*) Which was more unpleasant?
   (*c*) Can you tell me why?

RESULTS AND DISCUSSION

One *S* pressed the same lever on all 90 trials. His results were discarded, leaving a total of 20 *S*s.

In order to assess the effect of the shock variable on *S*s' responding, the total number of responses to the delay (D) lever was calculated for each *S* for Trials 31–60 (preshock trials) and for Trials 61–90 (shock trials). The difference between these measures was obtained for each *S*, and the (correlated) t test was applied to the 20 scores. The mean number of responses to the D lever for Trials 31–60 and for Trials 61–90 was 15.40 and 12.60, respectively. Although the mean difference, 2.80, was moderate in size, it proved to be reliably different from zero, $p < .01$. Thus, groupwise, pressing the lever that produced randomly delayed shock (RDS) was inhibited.

In answer to Question 4 of the questionnaire, 16 *S*s found RDS more unpleasant, two *S*s found IS more unpleasant, and for the remaining two *S*s both types of shock were equally unpleasant (Table 1). Applying the sign test to these data, we find that we may reject the hypothesis of equal unpleasantness of IS and RDS at the .01 level. Thus, both phenomenologically and with respect to its effect on an instrumental response, we have verification of the greater aversiveness of the unpredictable shock for the group as a whole.

There were, however, rather wide individual differences in the inhibiting effect of RDS. While 13 *S*s showed a decline in responding to the D lever during the shock trials, six demonstrated a reverse tendency and one *S* was unaffected. However, the largest decline obtained was 81%, while the largest increase was only 23%.

TABLE 1

SUMMARY OF THE SUBJECTS' RESPONSES

S	*S*s' Estimate of % of Trials D Lever Was Correct		*S*s' Estimate of % of Responses to D Lever		Actual % of Responses to D Lever		Type of Shock Found More Aversive
	Preshock	Shock	Preshock	Shock	Preshock	Shock	
1[a]	65	55	45	50	48	50	RDS
2	50	10	50	10	62	43	RDS
3	50	50	50	50	42	37	RDS
4[a]	40	55	56	50	57	66	RDS
5	50	60	50	25	43	27	RDS
6	50	50	50	15	55	30	RDS
7	50	45	50	25	55	30	RDS
8	40	80	60	20	55	37	RDS
9	50	25	50	20	43	50	RDS
10	40	50	40	40	50	47	RDS
11	50	50	65	75	53	53	RDS
12[a]	65	50	60	75	53	60	RDS
13	43	20	50	10	53	37	RDS
14	50	25	50	10	53	27	RDS
15	50	50	40	40	40	30	Equal
16	50	50	50	50	57	50	IS
17	40	50	60	50	57	43	RDS
18	60	60	50	80	50	53	IS
19	35	20	30	40	47	60	Equal
20	40	70	50	5	55	10	RDS
Means	48.4	46.3	50.3	37.0	51.4	42.0	

[a] Confusion as to D lever.

An interesting finding was the marked degree of impairment of cognitive processes which occurred during the shock trials. This impairment is illustrated by a comparison of the errors associated with Questions 1*a* ("Before shock was introduced, on what percentage of the trials did the left light go on?") and 1*b* ("After shock was introduced?"). We calculated for each *S* the absolute error involved in each of his judgments and found that while the mean absolute error for Question 1*a* was 5.6, it was more than twice as large for Question 1*b*, viz., 11.8. This increase in error associated with the shock trials is not quite significant by the two-tail Wilcoxon matched-pairs signed ranks test ($p \approx .07$). (Note that three *S*s answered Question 3 incorrectly.)

Pursuing the point further, we examined the results of Questions 2*a* ("Before shock was introduced, on what percentage of the trials did you press the left lever?") and 2*b* ("After shock was introduced?"), and once again found a decline in the accuracy of judgments made for the shock trials. The mean absolute error for the preshock trials was 6.0 and rose to 14.4 for the shock trials. Applying the previous test, we find that the difference is significant at less than the .01 level.

Thus, the conclusion seems warranted that something present during the shock trials disrupted the cognitive processes necessary for subsequent accurate recall. It is important to note that both recall of external stimulus events and recall of *S*'s own behavior suffer about equally from such disruption. This fact argues against a "distraction" interpretation, in the sense of failure of the stimulus to register due to faulty receptor orientation. Clearly, such an interpretation could not apply to recall of *S*'s own behavior.

Unfortunately, we are unable to determine the relative contribution of RDS and IS to the decline in accuracy of recall, a question of considerable interest. We also cannot evaluate the factor of unequal samples (60 preshock trials versus 30 shock trials). But in view of the fact that the shock trials had the benefit of recency and considering some of the enormous errors involved, it is doubtful that this factor can be of great significance.

Closer inspection of the data revealed some interesting patterns which, because of the ad hoc nature of their emergence and the small number of *S*s involved, can only be suggestive. We wish, nevertheless, to present them at this time for their heuristic value, inasmuch as we do not plan to return to this problem.

For the purposes of the following analysis, *S*'s instrumental responses and his answers to the questionnaire were converted from left and right responses to responses to the D and the I lever. Thus, for example, the figures reported in Columns 4 and 5 (Table 1) were obtained from the answers to Questions 2*a* and 2*b*. Similarly, the figures in Columns 2 and 3 are based on the answers to Questions 1*a* and 1*b*. The items of the questionnaire were formulated so as to be unambiguous and as unsuggestive to *S* as possible. In all that follows, the translation from the basic data must be kept in mind, since it represents an unassessable factor.

Pattern A is illustrated by the results of *S* 9. This *S* estimated that he pressed the D lever on 20% of the shock trials and claimed that the D lever was correct on 25% of these trials, whereas of course the D lever was correct on 50% of the shock trials. One might be inclined to ascribe this gross error in estimating how frequently the D lever was correct to nothing more than *S* "rationalizing" his behavior. However, the fact is that *S* actually pressed the D lever on 50% of the shock trials (an *increase* over his rate of pressing this lever during the preshock trials). Note that similar results obtain for *S*s 2, 13, and 14, though to a lesser extent. In fact, if we consider the five *S*s who estimated the D lever to be correct on 25% or fewer of the shock trials (note that the next lowest estimate is 45%), these *S*s (2, 9, 13, 14, and 20) turn out to markedly underestimate the frequency with which they pressed the D lever during the shock trials. Their average error (the error obtained by subtracting the estimated percentage of responses to the D lever during the shock trials—Column 5—from the actual percentage—Column 7) is plus 22.4, whereas the average error for the entire group is only plus 5. These *S*s, as a matter of fact, yield five of the six most extreme positive errors. One might speculate that the four *S*s named found the RDS much more aversive than the IS and in part resisted its effect on the instrumental response, but only at the cost of severe distortion of recall of both their behavior during the shock trials and the stimulus events thereof.

Pattern B seems to depict a more "adaptive" approach to the experimental situation. These *S*s find the RDS more aversive than the IS and protect themselves by avoiding it. Further, they recognize their avoidance and in general maintain good perception and recall of their behavior and environment. This pattern is well illustrated by *S* 7. Note the avoidance of the D lever during the shock trials and the accuracy of this *S*'s estimates. *S* 5, and perhaps to a lesser extent 6, also display this general pattern.

Pattern C seems rather prevalent, namely, *S* shows fair to good accuracy in his estimates, claims his behavior was the same in the shock and preshock trials, but actually shows a slight avoidance of the D lever during the shock trials. *S* 16 typifies this pattern, also displayed by *S*s 3, 10, and 15. A reasonable interpretation is that the RDS is only slightly more aversive to these *S*s, if at all, and its effect on the instrumental response is too slight either to be noticed or recalled. Note that only two of these four *S*s admit to the RDS being more unpleasant.

Other behavior worthy of note is *S* 18's claim that the IS was more unpleasant, his unaffected instrumental behavior, and the abnormally high estimate of the frequency of his D lever presses during the shock trials.

In conclusion, the authors were impressed by the magnitude and variety of effects observed under the experimental variable, particularly in view of the unselected and homogeneous sample employed. We view the above experimental situation as potentially fruitful for the study of how differential susceptibility to the effects of RDS and particular patterns of behavior, as illustrated above, are related to personality dimensions.

SUMMARY

Twenty *S*s were required on 90 trials to guess whether a right or a left light would go on, indicating their guess by pressing a lever below the appropriate light. Trials 1–60 were nonshock trials. During Trials 61–90, *S* received an immediate shock (IS) every time he pressed one lever and a randomly delayed shock (RDS) every time he pressed the other lever, irrespective of whether he guessed correctly or incorrectly. Comparison of *S*s' behavior on the shock and nonshock trials revealed:

1. A significant avoidance of the lever leading to RDS;
2. A significantly greater number of *S*s judged the RDS as more unpleasant than the IS;
3. A marked impairment of *S*s' recall of stimulus events and their behavior during the shock trials;
4. Several suggestive patterns of adjustment to the experimental situation.

REFERENCES

SIDMAN, M., & BOREN, J. J. The relative aversiveness of warning signal and shock in an avoidance situation. *J. abnorm. soc. Psychol.*, 1957, **55**, 339–344.

(Received November 7, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

INCIDENTAL STIMULATION:

A STUDY OF PRECONSCIOUS TRANSFORMATIONS[1]

FRED PINE

Research Center for Mental Health, New York University

THE relevance of a stimulus to an individual's ongoing adaptive functioning is likely to be a major determinant of the nature and course of its effects. The present study deals with the participation of incidental and focal stimuli in stories that persons produce immediately following stimulus exposure. Differences in such participation related specifically to the incidental or focal character of the stimuli are explored; the main concern is with qualitative transformations undergone by incidental stimuli.

In attempting to account for the transformation from latent meaning to manifest content in dreams, Freud (1900) described a kind of thinking with "rules"—condensation, displacement, symbolization—quite out of accord with the rules of formal logic and with the experience of conscious waking thought. This mode of thinking, characterized by the "primary process," emerges in normal adult life in conditions where the reality-adaptive functions of the ego, including the ego's function of critical guidance of the direction of thinking, are in abeyance—for example, in dreams or on the analytic couch. The central hypothesis of this study is that the vicissitudes of irrelevant incidental stimuli are governed more by the laws of primary process thinking while those of relevant focal stimuli are governed more by the laws of the secondary process. Stimuli that are irrelevant to reality-adaptation may escape being used in thought that is directed towards adaptive manipulation of reality (the "secondary process") and may undergo transformations having no reality-adaptive significance, reflecting the archaic forms and highly idiosyncratic drive-related content of primary process thinking.

The present study is closely related, by way of contrast, to studies of "discrimination without awareness" (see Adams, 1957); also McConnell, Cutler, & McNeil, 1958). The focus of these studies is on what may be described as the "accuracies" in behavior with respect to a stimulus presented below the threshold of awareness. In contrast, the present study deals with the "inaccuracies" in the effects of marginal stimuli (subliminal or incidental)—effects traceable to a stimulus but nonetheless quite altered from the initial stimulus. It is also closely related to studies of the effects of subliminal stimulation which indicate that unreported stimuli emerge in cognitive productions subsequent to the stimulation (Allers & Teler, 1924; Bach & Klein, 1957; Fisher & Paul, 1959; Klein, Spence, Holt, & Gourevitch, 1958; Paul & Fisher, 1958; Poetzl, 1917; Shevrin & Luborsky, 1958; Smith, Spence & Klein, 1959). An additional observation in these studies has been that what eventually emerges in subsequent dreams or images is related to but certainly not always identical with the initial stimulus. The present study is focused specifically on these qualitative alterations in the stimulus (here, an incidental stimulus) and the theoretical implications of such stimulus transformation.

A common criticism of studies of discrimination without awareness and of subliminal stimulation has been that the stimuli were not truly subliminal and that subjects were in fact responding to some partial information stemming from the stimulus (see Adams, 1957; Bricker & Chapanis, 1953; McConnell et al., 1958). This critique has at times tended to direct theoretical discussion towards the issue of whether (and how) a truly subliminal stimulus can influence behavior. Even if the hypothesis of partial cues is helpful in explaining the influence of such stimuli, however, the hypothesis contains no direct suggestion for explaining the qualitative changes that they undergo. The use of incidental rather than subliminal stimuli in the present study, in addition to extending related research using subliminal stimuli (cf. Klein, 1958), was intended to bypass the issue of subliminal-or-not for the moment and to refocus on the issue of qualitative differences in the effects of stimuli when they are in focal awareness as compared with when they are incidental and irrelevant to reality-adaptation.

---

[1] The study is one of a series supported by a grant from NIMH (Project Code Number: M-782, Cognitive Processes Below the Level of Awareness). I am indebted to my colleague, George S. Klein, whose thinking had important influences throughout on the basic design and the theoretical orientation of the research. The staff of the Research Center for Mental Health contributed to the study by carrying out rating tasks and by reading the manuscript. I wish especially to thank Robert R. Holt for his many suggestions and Frances Lippmann for her major share in the rating procedures and the data analysis.

METHOD

*S*s were 24 male undergraduates seen individually in one-hour sessions. They were asked to invent two stories (the "pre" stories). Next, two stimuli were presented simultaneously: a "focal" one that *S* read and an "incidental" one, produced in an adjoining room, that was experienced by *S* as irrelevant "noise." The stimuli were verbal descriptions of a cow and of a hook; each was incidental for half of the *S*s. Following this, *S*s again were asked to invent two stories (the "post" stories). Data were analyzed by comparing influences of the "cow" and the "hook" stimuli on the poststories under the incidental and focal conditions.

The stimuli were written by *E* to suggest particular kinds of indirect associations. It was therefore possible to predict specifically at least some of the directions of change in the content of the stimuli when administered incidentally; emergence of a transformed derivative of a stimulus rather than some intact stimulus content was expected for incidental stimuli. Psychoanalytic theory and clinical description underly the thinking about the associative value of the stimuli.

The "cow" stimulus was written to convey oral-passive connotations, using the mother-infant relationship (feeding, contact, warmth) as a model. It reads as follows:

> The ordinary cow is an animal which is indispensable to mankind. Among the domesticated animals, it holds a place of special importance for us all. The dairy industry, and in turn all consumers, are dependent upon the cow as a basic source of supply. Cows thrive in the warm climates of the temperate zones, in the protective atmosphere of small farms. They are found especially on gently rolling meadowlands where serene warmth promotes a contented existence. Lands where grass is lush and plentiful are the special home of the cow. The cow itself is a warm-blooded animal. On contact, its flesh too has a moist warmth as though the inner warm-bloodedness pervades the whole being of the animal. It is a soft animal. At its fullest development, the cow has a fleshy covering which envelops the whole of its frame. The cow is warm; it is soft. Cows live in groups. They follow one another and hover close to each other. Even when spread out across a small meadow, there is a basic "oneness" among the several cows of a small herd—all living together in the lush and warm meadow which is their home.

Any of the manifest content of the stimulus appearing in the stories was considered a relatively direct stimulus derivative. Any stray content that seemed more distantly related to the stimulus, either on theoretical grounds or on the basis of clinical impression, was considered a transformed derivative. Thus, direct derivatives could include references to cows, grass, and warm climates. Transformed derivatives could include oral-passive themes such as mothering and dependent relationships.

The "hook" stimulus was written to convey phallic-aggressive connotations, using as a model the syndrome of masculinity, aggression, intrusiveness, voyeurism, and exhibitionism which psychoanalytic writers have described as characteristic of the phallic stage of development. It reads as follows:

> If your glance has ever been captured by the silvery glint of a steel framed hook, you may have sensed its enormous potential for mechanical action. The capacity of a hook to pierce the surface of huge bulks makes it the tool of choice for use when heavy, active, and effective work must be carried out. Hooks can be moved rapidly, but there is a firm sureness to their movements when they are properly used. They hiss through the air and seize firmly upon their targets. The larger hooks can be used to raise great weights off the ground, to swing them here and there at will, to control the movements of other objects with a steady and sure grasp. In viewing the hook, one cannot avoid being aware of its metallic hardness, its steel grey coldness, and of the stark prong at the end of the molded curve which forms the hook. The hard steel is an essential part of its makeup. The hook's unbending arc is a bulwark of the power required for the immense tasks it must perform. But the effectiveness of the cold steel is ultimately determined by the sharpness of the polished edge. Coldness, hardness, silvery steel: these are the qualities of the hook.

Direct derivatives of the hook stimulus could include references to sharpness, mechanical work, swift motion, and the like. Transformed derivatives could include, for example, aggressive, voyeuristic, or exhibitionistic themes.

The procedure, step-by-step, was as follows. *S* and *E* were alone in an office that was sufficiently isolated to virtually exclude extraneous noise. *S* had been told only that the study had to do with "imagination and some other things." To begin with, *S* was given some material from the MAPS Test (Shneidman, 1951). A card with a picture of an empty theater stage was used along with four blank-faced figures usually seen as boy, girl, man, and woman. *S*'s task was to build a scene on the stage using any or all of the four figures and then to tell a story about it. *S* was then given the blank TAT card and was asked to tell another story. Standard TAT instructions were given for both stories; instructions were written out and given to *S* to eliminate any need for a potentially biasing inquiry. All stories were tape recorded and later transcribed.

Following the storytelling, *S* was told that he would get a rest from imagination tasks while *E* got a measure of his "ordinary reading rate" in reading "just for comprehension." Instructions were for *S* to read a passage silently several times, announcing each time a reading was completed so that *E* could note the time. *S* was encouraged to read at a steady rate even as he got increasingly familiar with the material after the first reading; he was told that the aim of the study was to examine "consistency" in reading rate. This instruction succeeded in keeping *S*s well focused on the reading task even though the incidental stimulus was audible from an adjoining office simultaneously. *S* was given either the cow or the hook passage. *E* had *S* read this over and over, at least until the broadcast of the incidental stimulus was completed; 18 *S*s read the focal stimulus just three times and the remaining 6 *S*s read it two to five times.[2] After the reading, *E* reviewed the salient points in the passage for *S* in a standardized manner under the guise of making certain *S* clearly understood them; the intention was to maximize the focal quality of the passage.

Ten seconds after *S* began reading the focal stimulus, *E* (unbeknown to *S*) signaled an assistant in an adjoining room to broadcast the incidental stimulus. The incidental stimulus (cow or hook) was the one other than the one *S* was reading. It was tape recorded by a male voice (not *E*'s) and always played at the same volume. The voice volume was such that it could be heard in the experimental room as clearly as though it were being read at close to ordinary speaking volume in the same room as *S*. The incidental stimulus took just under two minutes to read; it was read once. The subgroups of the study are hereafter referred to by the incidental stimulus they received; there were 12 *S*s each in the "cow-incidental" and in the "hook-incidental" groups.

Various devices were used to increase the likelihood that *S* would experience the incidental stimulus as irrelevant. When *E* and *S* first arrived at the experimental room, the door was always "accidentally" locked and *E* had to get the key from someone who was "working" next door. Then, during *S*'s telling of the pre-experimental stories, the phrase "testing—1, 2, 3" was broadcast several times on two separate occasions. Finally, the incidental stimulus itself was preceded by a reading of "testing—1, 2, 3" five times; it was found in discussion with pretest *S*s that this gave *S* a chance to experience the interruption to his own (focal) reading and to accustom himself to the intrusion even before the incidental stimulus itself was broadcast.

Following the reading task, *S* was told that he would now "return to the imagination experiment" and was instructed to tell stories once again. The same MAPS Test material was given, followed by the blank TAT card. Story instructions for the poststories were identical with those for the prestories except that *S* was told to tell a different story from the one he had told the first time.

Memory for the focal and incidental stimuli was next investigated. *S* was asked to tell all he could recall of the material he had read. He was then given a recognition test consisting of 42 phrases, half of which had appeared in the written passage; *S* checked off those he recognized. Finally, *S* was reminded that he "might remember that there was some interfering noise while he read the passage earlier. He was then given the same recall and recognition tests for the incidental stimulus.

Generally, *S*s were quite surprised when they were asked about the incidental stimulus. Only one *S* recalled more than one or two important words (e.g., cow, skin, metal, silvery), although five *S*s were able to recognize correctly from three to seven phrases. Only one *S* said that he thought that the incidental stimulus might have something to do with him, and no *S*s suspected any connection between the storytelling and the incidental and focal stimuli. The incidental stimulus was typically experienced as an interference, sometimes with annoyance, and attempts were made to ignore it. Memory for the incidental stimulus was thus quite poor even though it was well above the auditory threshold. More significant from the point of view of the study, however, is the fact that *S*s experienced the incidental stimulus as irrelevant.

RESULTS

Several analyses of the stories were undertaken, each designed to test the central hypothesis that irrelevant incidental stimuli, in contrast to relevant focal stimuli, tend to undergo transformations reflecting the archaic forms and drive-related content of primary process thinking.

Quality of main relationship. Story content that could potentially reflect the emergence of transformed derivatives of either the incidental or the focal stimulus was first examined. It was expected that such story content would show the effects of the incidental rather than the focal stimulus. The quality of the relationships between the characters in the stories was analyzed in accord with this expectation.

Interpersonal relationships were rated for their oral-passive and phallic-aggressive qualities. Such drive-related qualities as these in relationships between people are related to the peripheral connotations of the cow and hook stimuli, respectively, but are only indirectly related to the manifest stimulus content; they were conceived of as transformed derivatives of the cow and hook stimuli. It was predicted, therefore, that oral-passive and phallic-aggressive qualities in interpersonal relationships would be influenced by incidental rather than focal stimuli.

The interpersonal relationships in the stories were analyzed in three parallel ways. In each, a continuum was described ranging from categories with connections primarily with the

---

[2] The data for these six *S*s suggests that the more the focal stimulus was read, the less was the effect of the incidental stimulus.

(transformed) cow stimulus through to those with connections primarily with the (transformed) hook stimulus. The categories are given here along with the score (in parentheses) assigned to each; the numerically higher scores were always assigned to those categories conceived to be related to the cow stimulus. In the first analysis, after Erikson (1950), relationships were categorized as "receiving" (3), "giving" (2), "taking" or "withholding" (1), and "intruding" (0). In the second analysis, after Horney (1937), relationships were categorized as "moving towards" (2), "moving away from" (1), and "moving against" (0). In the third analysis, relationships were categorized as "positive" (2), "mixed" (1), and "negative" (0). These three systems of analysis are obviously not independent; for example, a "positive" relationship is likely to be "receiving" or "giving" and unlikely to be "moving against." The results give different ways of looking at one over-all finding.[3]

The placement of stories into categories was carried out by a rater who knew neither which stories were "pre" and which "post" nor which stimulus was incidental and which focal. For each of the three systems of analysis, two scores were derived for each *S*: a "post" score derived by adding up the category scores for the two post stories, and a "difference" score reflecting the change in score from the two pre- to the two poststories. The 24 *S*s were ranked on each of the three categorizations, and differences between the cow-incidental and hook-incidental subgroups were tested by the Mann-Whitney test. Results are presented in Table 1.

TABLE 1

Signifcance of Differences between Ranks of Cow-Incidental and Hook-Incidental Groups on Interpersonal Relationships Categories

Rating Dimension	"Post" scores		"Difference" Scores	
	z Score[a]	p Value*	z Score[a]	p Value*
Receiving–Intruding	2.80	.01	1.82	.05
Towards–Against	1.01	.16	1.36	.09
Positive–Negative	1.39	.09	1.85	.05

[a] All differences are in the predicted direction.

* One-tailed tests.

All results were in the expected direction. Findings were strongest for the receiving–intruding dimension where a finer categorization of relationships in the stories made for a wider range of scores. Thus, stories of *S*s in the cow-incidental group tended to involve relationships at the "receiving" end of the continuum, while stories of *S*s in the hook-incidental group tended to involve relationships at the "intruding" end. An examination of the category analyses from another viewpoint indicated further that, on all three systems of categorization, *S*s in the cow-incidental group showed an absolute increase of cow-related interpersonal relationships in the post- over the prestories, while *S*s in the hook-incidental group showed an absolute increase of hook-related interpersonal relationships.

These results are especially striking since all *S*s received both the cow and the hook stimuli. Furthermore, *S* read the focal stimulus several times in contrast to hearing the incidental stimulus only once. In spite of this, or as it will later be argued, *because* of this, the effects just described were related to the incidental stimulus and went in a direction opposite to the focal stimulus.

Cast of characters. The kinds of characters used in the stories were also examined to study the effects of the experimental stimuli. The oral-passive connotations of the cow stimulus could produce a tendency towards the use of children as characters in the subsequently produced stories. An effect of this kind is only indirectly related to the manifest content of the cow stimulus and was conceived of as a transformed stimulus derivative; as such, these effects were expected to show the influence of incidental rather than focal stimuli. It was predicted, therefore, that the oral-passive connotations of the cow stimulus would make for greater use of children in stories of *S*s in the cow-incidental group in contrast to the greater use of adults in stories of *S*s in the hook-incidental group.

---

[3] These abstract category titles may give a misleading impression regarding how directly or indirectly the interpersonal relationships are related to the manifest content of the cow and hook stimuli. For example, "moving towards" was typically represented in the stories not by people standing near each other (a more direct derivative) but in a wide range of idiosyncratic ways, quite unrelated to the manifest content of the cow stimulus—for example, a man buying his son a new car or a family making plans for a vacation.

The presence of men, women, boys, and girls was tabulated for all stories; only one tabulation was made for each of these four types in each story regardless of the number of any one type (because of the frequent use of groups of people as story characters where the number in the group was unspecified). The ratio of child types (boys or girls) to the total number of types used in the two poststories, and also the difference in this ratio from pre- to poststories, was computed for all *S*s. Scores were ranked and the Mann-Whitney test (one-tailed) was used to test the significance of the differences between ranks in the two subgroups. Results were in accord with initial expectations. The use of children in the poststories was significantly greater in the cow-incidental group ($z = 2.02$; $p = .025$); the converse of this regarding the greater use of adults by *S*s in the hook-incidental group is also true. The same trend also appeared in the pre-to-post difference scores ($z = 1.44$; $p = .075$). Absolute changes were in accord with the relative rank differences; in the cow-incidental group there was an increase in the ratio of child types in the two poststories as compared to the two prestories, and in the hook-incidental group there was an increase in the use of adult types.

In a further analysis, a rater who did not know which stimulus was incidental and which focal judged whether a child or adult was used as the single principal character in each of the poststories. In this case, since no decision could be made for several of the stories, the scores for use of child or adult as the main character are not directly complementary to one another. Separate rankings were made for all *S*s of the number of poststories in which a child was used as the main character and also for the the number which used an adult. The Mann-Whitney test was again used to test for differences. More children were used as main characters by *S*s in the cow-incidental group ($z = 1.65$; $p = .05$), and more adults by *S*s in the hook-incidental group ($z = 1.53$; $p = .07$). This result supports the one reported above for ratio of children and adults used, but is clearly not independent of it. The two analyses give different ways of looking at a single finding. However, these child-adult findings are conceptually distinct from the earlier reported quality of main relationship findings; relationships involving either children or adults could be either, for example, positive, mixed, or negative. Both of these over-all sets of findings demonstrate, from one point of view, the emergence of the incidental stimuli in transformed ways and, from another point of view, the greater effect of incidental than of focal stimuli on aspects of the stories which are only indirectly related to the manifest stimulus content.

The data on the cast of characters of stories were also analyzed in terms of male and female, but no significant differences between subgroups were obtained. It had been expected that the use of male characters, a transformed derivative of the (phallic-aggressive) hook stimulus, would be more prevalent in the hook-incidental group while the use of female characters would be more prevalent in the cow-incidental group. One may speculate that in a male sample with strong tendencies to use males as identification figures in the stories, an effect of variations in immediate stimulus conditions on sex of story characters cannot readily emerge.

Over-all incidental-focal judgments. A global analysis of the story content was also undertaken. The stories were submitted to a group of raters who were asked to judge which stimulus had been incidental and which focal. Each of the 24 protocols of four stories was judged by three different raters. One rater judged all 24 protocols, while the second and third raters varied from protocol to protocol; 15 raters were used in all. Raters were instructed to make their judgments by using the hypothesis that the incidental stimulus would emerge in the stories in more transformed ways and the focal in more direct ways. Raters were informed which two stories were "pre" and which two were "post"; they were advised to use the prestories as a base line for revealing characteristics of the stories attributable to personality factors in *S* rather than to the stimulus.

Kinds of story content which raters singled out as stimulus derivatives and used in determining the final judgments may be illustrated as follows: Items conceived of as relatively direct derivatives of the cow stimulus included phrases such as "nice countryside," "playing in a field," "at the market," "average," and "foodstuffs," as well as themes such as life on the countryside and animal stories—all of which can be traced in a careful reading of the cow stimulus. Relatively transformed derivatives of the cow stimulus included phrases like "holding his stomach," "gaining sympathy," and "at the dining room," as well as themes such as a mother protecting her son, and receiving welfare benefits from the state. The rationale for the selection of transformed derivatives varied with the judgment of the individual rater; in the examples just given, the oral-passive quality of the derivatives was a determining factor. Relatively direct derivatives of the hook stimulus included such phrases as "hurrying," "working hard," "glancing around him," and "effective work" as well as themes such as repairing things and going to engineering school. Relatively transformed derivatives of the hook stimulus included phrases such as "becomes successful" and "a cold snowstorm" as well as themes of aggression, death, voyeurism, and exhibitionism.

Whenever at least two of the three raters judged correctly which stimulus had been incidental for an *S*, the combined judgment was considered correct. (Since one judgment was missing for one *S* and the remaining two judges split, this analysis is based on an N of 23.) Of 23 *S*s, 17 were judged correctly ($p =$ .02, Fisher exact test). Judges were able to do equally well for both the cow-incidental and the hook-incidental subgroups. Nine of 12 judged correctly in the cow-incidental group ($p = .07$) and 8 of 11 in the hook-incidental group ($p = .11$); smaller sample size here accounts for the failure to reach the .05 level.[4]

DISCUSSION

Qualitative differences thus appeared in the effects of relevant focal stimuli and of irrelevant incidental stimuli on subsequently produced stories. Incidental stimuli tended to emerge in transformed ways in the stories. Were one to look only for intact, rather than transformed, derivatives of incidental stimuli, some of the major effects of such stimuli would be obscured. A recent study (Goldstein, 1958) in which direct measures of the registration of subliminal stimuli proved less "sensitive" than indirect measures supports the same view. It was suggested, regarding the central hypothesis of the present study, that results could be explained by postulating that different streams of thought intervene between stimulus exposure and the emergence of effects in conditions of incidental and focal stimulation. Psychoanalytic theory (cf. Freud, 1900; also Rapaport, 1951) gives the fullest description of such contrasting thought streams, subsumed under the concepts of "primary process" and "secondary process."

The focal stimuli of the present study fall under the sway of adaptive intentions. An *S* instructed to read for comprehension functions adaptively by trying to grasp the literal content of the cow or the hook passage. When recall of the focal stimulus is demanded, adaptation and communication are only served by reporting on the cow as a cow, on the hook as a hook—and nothing else. It is precisely this direction of thinking by adaptive intentions, including the intentions to communicate effectively and to cope adequately with the requirements of external reality, which defines the secondary process. It may be suggested further that once a focal stimulus has achieved a clear-cut definition in full awareness it does not as readily reappear in distorted ways in imaginal productions; such a process would parallel the common experience wherein, once structure has been given to a vague stimulus, it can no longer be seen readily in any other way.

In contrast to the focal stimuli, the incidental stimuli were irrelevant to the *S*s' tasks; they appear to have entered only minimally into awareness and adaptive thinking. It is unlikely, in this setting, that the cow and hook stimuli attained a fixed and literal identity comparable to the clarity and definiteness of concepts in secondary process; the transformed quality of the derivatives reflect a fluctuating and many-sided definition of the incidental stimuli. By using as a model both the drive-directedness and the nonrational forms of primary process thinking, we were able to predict some of the specific effects of the incidental stimuli. It may be that the peripheral oral-passive and phallic-aggressive connotations of the stimuli activate drives in the person that tend towards discharge by a "spilling over" of drive-related content into the stories. In addition, it may be that the incidental stimulus has the quality of an incompleted task (cf. Zeigarnik, 1927) which, never having become fully structured in awareness, undergoes a continuing and relatively autonomous process of transformation within the person (cf. Rapaport, 1957). Thus, what emerges in the story protocols as transformed derivatives of incidental stimuli may actually be the end result of some associative process. The incidental stimuli set off a chain of association to other material relevant to specific drives which have been activated and along pathways with such nonrational formal links as negation, clang association, and symbolization.

---

[4] Two additional analyses of the data were carried out using some further judgments made by the raters who gave the incidental-focal judgments. These involved a listing of all recognizable derivatives of either the cow or the hook passages in the stories, and ratings of the degree of orality and of aggression of the stories. Scores derived here proved to be unreliable. However, statistical analysis was carried out with these scores and did not confirm the hypothesis that incidental stimuli emerge in more transformed ways.

What are the conditions for a stimulus to lead to such associations and thus to emerge only in transformed ways in later thinking? At least two are suggested.

First, when a stimulus is irrelevant for adaptive functioning the likelihood of its entering into the meandering stream of primary process thought is increased. Incidentalness of a stimulus may at times be associated with irrelevance, for example as the incidental stimuli were in the present study, but it is by no means true that all incidental stimuli are irrelevant for adaptive functioning. Whether decreased awareness and decreased relevance produce effects similar in all respects to one another is unlikely but that at least some of their effects are similar is suggested by the correspondence of the results of the present study with those of studies of subliminal stimulation. The subliminal stimulus (see Fisher & Paul, 1959; Shevrin & Luborsky, 1958) does not generally tend to produce direct representations of itself (as advertisers would like to have it), but instead emerges in quite transformed ways.

The use of imaginative story production, and probably of fantasy in general, as a vehicle to carry the emerging derivatives of the stimulus also heightens the likelihood of stimulus transformation. Fantasy, involving some relaxation in customary standards of reality orientation, throws open the doors to those thought contents which are ordinarily kept out of awareness because of their irrelevance and also permits breakdowns in logic that would not be acceptable in reality-adaptive thinking. It is probably because of the use of TAT type stories to express the stimulus derivatives that even the focal stimuli emerged with some transformation in the poststories. In contrast, memory for the focal stimuli showed almost no transformation of the stimulus content.

Even though this study employed incidental stimuli and imaginative productions, both of which facilitate emergence of derivatives suggestive of nonrational thought processes, it should be noted that the effects of the incidental stimuli were still limited. The derivatives of the incidental stimuli did not completely "capture" the themes of the subsequent fantasies. Nor did the presence of incidental stimulation markedly lessen the effectiveness of memory for focal stimuli. The special status of an irrelevant incidental stimulus gives it an advantage for certain kinds of effect but not a power transcending the reality-oriented and adaptive functioning of the ego based upon intake of relevant stimuli in full awareness and acting in lines guided by conscious intentions.

SUMMARY

The present study of effects of incidental stimulation on subsequently produced stories, which contrasted the quality of effects produced by incidental as opposed to focal stimulation, is closely related to recent studies on the effects of subliminal stimulation.

Twenty-four male *S*s were seen individually for a study of "imagination." Each began by telling two stories, with standard TAT instructions, to ambiguous stimuli. Following this, two stimuli were administered simultaneously. The "focal" stimulus was a written passage that *S* was asked to read as a measure of his "ordinary reading speed." The "incidental" stimulus was a similar passage read by a male voice and broadcast through the wall from an adjoining room; this stimulus was consistently experienced as irrelevant "noise" by *S*s. The two passages, describing a cow and a hook, were respectively devised to convey oral-passive and phallic-aggressive connotations; each was incidental half of the time (while the other was simultaneously focal). Finally, *S* again told two stories.

Contrasting effects of the incidental and focal stimuli were sought in the poststories, using prestories as a control. The prediction that irrelevant incidental stimuli tend to emerge in the stories in more transformed ways, and relevant focal stimuli in more direct ways, was tested; this prediction was derived from the theoretical hypothesis that irrelevant stimuli enter into primary process modes of thinking more than do focal stimuli. The results gave general support to the prediction. For example, blind raters were able to judge which stimulus had been incidental and which focal by basing judgments on the prediction that incidental stimuli would tend to emerge in transformed ways.

The relation of the study to recent work on subliminal stimulation, the relation of the results to psychoanalytic theory of primary and secondary process, and the conditions for transformation of incidental stimuli were discussed.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

ALLERS, R., & TELER, J. Uber die Verwertung unbemerkter Eindrucke bei Associationen. *Z. ges. Neurol. Psychiat.*, 1924, **89**, 492–513.

BACH, S., & KLEIN, G. S. The effects of prolonged subliminal exposures of words. *Amer. Psychologist*, 1957, **12**, 397–398.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.

FISHER, C., & PAUL, I. H. The effect of subliminal stimulation on images and dreams: A validation study. *J. Amer. Psychoanal. Ass.*, 1959, **7**, 35–83.

FREUD, S. *The interpretation of dreams*. (Original German ed., 1900. New York: Basic Books, 1955.)

GOLDSTEIN, M. Subliminal stimulation: Its cue value in fantasy. Paper read at American Psychological Association, Washington, D. C., September 1958.

HORNEY, KAREN. *The neurotic personality of our time*. New York: Norton, 1937.

KLEIN, G. S. On subliminal activation. Paper read at American Psychological Association, Washington, D. C., September 1958.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNA. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

MCCONNELL, J. V., CUTLER, R. L., & MCNEIL, E. B. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, **13**, 229–242.

PAUL, I. H., & FISHER, C. The effects of subliminal visual stimulation on images and dreams. Paper read at American Psychological Association, Washington, D. C., September 1958.

POETZL, O. Experimentell erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Z. ges. Neurol. Psychiat.*, 1917, **37**, 278–349.

RAPAPORT, D. Towards a theory of thinking. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.

RAPAPORT, D. Cognitive structures. In J. S. Bruner, E. Brunswik, L. Festinger, F. Heider, K. F. Muenzinger, C. E. Osgood, & D. Rapaport, *Contemporary approaches to cognition*. Cambridge: Harvard Univer. Press, 1957.

SHEVRIN, J., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.

SHNEIDMAN, E. S. *Thematic test analysis*. New York: Grune & Stratton, 1951.

SMITH, G., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59**, 167–176.

ZEIGARNIK, B. Uber das Behalten von erledigten und unerledigten Handlungen. *Psychol. Forsch.*, 1927, **9**, 1–85.

(Received November 14, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

PARENTAL IDENTIFICATION, ACCEPTANCE OF AUTHORITY, AND WITHIN-SEX DIFFERENCES IN COGNITIVE BEHAVIOR[1]

JAMES BIERI
The New York School of Social Work, Columbia University

SEX differences in behavior continue to be reported in the literature of personality research, a recent example being the studies of anxiety reviewed by Jensen (1958). In the area of cognitive behavior, a rather consistent finding has been that males do better than females on certain perceptual tasks such as the embedded figures test (EFT) (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954; Bieri, Bradburn, & Galinsky, 1958). One way to conceptualize the basis of these sex differences is to analyze behavioral differences *within* each sex in terms of personality variables which, while common to both sexes, may be considered to be closely related to the sex role of the individual. Further, these personality variables should be logically related to the dispositions considered to be crucial in producing differential performance on the perceptual task being studied.

Witkin et al. (1954) consider differences in EFT performance to be due to field dependent-field independent orientations of the individual. Such a view, in turn, may readily be related to sex-role differences; that is, males are more field independent and females more field dependent. Presumably, in our culture, the male sex role involves more active, assertive, independent behaviors than does the female sex role. In addition to Witkin's own work, research by others at least partially supports this view. Thus, Marlowe (1958) has recently reported a significant negative relationship between succorance (dependency) and EFT performance.

These considerations suggest the value of analyzing within-sex differences in EFT performance as a function of two personality variables which conceptually should be related to field dependence-independence. The first of these is parental identification. It is assumed children learn behaviors more similar to those of the same sex parent. Whether this identification is achieved primarily through defensive strategies, as postulated in psychoanalytic theory, or whether it represents a more positive emulation of parental behavior, is an issue not directly of concern here. What is crucial is the assumption that a person's pattern of parental identifications is closely associated with his sex role behavior, and that these identifications should thus prove to be related to sex differences in cognitive behavior.

Assume that an individual develops identifications with both parents, but a stronger identification may exist with one parent than with the other. Assume further that characteristic sex role behaviors are associated with paternal and maternal identifications. Specifically, persons who identify with the father, in contrast to mother identifiers, should manifest greater independence, assertive and active coping behaviors, a syndrome resembling the field-independent person described by Witkin et al., (1954). In comparison, *S*s who identify primarily with the mother are assumed to manifest more field-dependent behaviors. These differential behaviors should apply to both sexes, e.g., women who identify with the father should be more field independent than women who identify with the mother.

The second personality variable hypothesized to be related to cognitive behavior is acceptance of authority (AA), one aspect of the broader authoritarian syndrome, akin to authoritarian submission as discussed by the California group (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). If AA reflects a more passive, conforming, and accepting orientation toward authority, such an orientation should be reflected in poorer or field-dependent behavior on the EFT.

These formulations lead to the following hypotheses:

---

[1] This study was supported by a research grant (M-1976) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

The writer wishes to acknowledge the assistance of Robin Lobeck in analyzing the data for this study.

1. Within each sex, and in both sexes combined, *S*s who identify more with the father score lower on the EFT than mother identifiers.
2. Within each sex, and in both sexes combined, *S*s who are low in AA score lower on the EFT than *S*s who are high in AA.
3. Within each sex, and in both sexes combined, *S*s who have low AA scores and who identify with the father score lower on the EFT than *S*s with high AA scores who identify with the mother.

METHOD

The *S*s consisted of 30 college undergraduates of each sex attending a large Eastern private university. Each *S* was administered the following tests in an individual session.

Embedded Figures Test. An eight-item form of Witkin's EFT was administered in standard fashion to each *S*. Previous research has demonstrated the equivalence of this form to a longer version of the EFT (Bieri et al., 1958). Since the EFT is scored for latency of response, a low score indicates better performance.

Parental identification. In a previous paper (Bieri, Lobeck & Galinsky, 1959) two aspects of the identification process were analyzed, namely, perceived similarity to a parent and degree of involvement with a parent. It was found that the measures of identification in terms of perceived similarity (interview and semantic differential) were more consistently interrelated for both sexes than were the involvement measures (interview and TAT). For this reason, the perceived similarity aspect of identification was focused upon in the current research, and the semantic differential was selected as the measuring instrument. The form used consisted of 12 bipolar adjectives that were rated on a seven-point scale. Each *S* first rated himself, then his mother, and then his father. Discrepancy (*D*) scores were calculated on each of the 12 scales, reflecting the disparity between *S*'s rating of himself and his rating of each parent. It was considered that *S* identified most strongly with that parent who had the lower *D* score.

Acceptance of authority. This variable was measured by a scale containing 20 items with high loadings on a factor called acceptance of authority by Bales & Couch (1956). The items have a mean loading of .65 on the AA factor, with a range from .36 to .76. In a previous report, where this scale is presented in detail, it was demonstrated that this scale correlates highly with the F scale (Bieri & Lobeck, 1959).

RESULTS

Contrary to expectation, the mean performance scores of the males and females on the EFT were not significantly different, the values being 32.31 and 33.03, respectively. In an earlier study (Bieri et al., 1958), using the same form of the EFT on samples of 50 males and 60 females, mean EFT scores of 27.90 and 41.04 respectively for each sex were obtained. In comparison to these earlier samples, the males in the present research thus appear to be more field dependent and, the females more field independent. Features of the present *S*s' parental identification to be discussed may have operated to produce this similarity between the sexes on the EFT.

TABLE 1

MEAN EFT SCORES OF *S*s GROUPED BY PARENTAL IDENTIFICATION AND ACCEPTANCE OF AUTHORITY

	Male	p	Female	p	Combined	p
Mother Ident.	36.00 (N = 13)	—	40.09 (N = 11)	.04	37.88 (N = 24)	.09
Father Ident.	27.47 (N = 15)		29.78 (N = 18)		28.73 (N = 33)	
High AA	42.00 (N = 10)	.02	37.30 (N = 10)	—	39.65 (N = 20)	.04
Low AA	21.33 (N = 9)		34.20 (N = 10)		28.10 (N = 19)	
High AA—Mother Ident.	67.25 (N = 4)	.06	44.00 (N = 4)	.10	55.62 (N = 8)	.02
Low AA—Father Ident	28.00 (N = 3)		27.00 (N = 4)		27.43 (N = 7)	

Note.—Mann-Whitney test; one-tail p values. One male in the low AA group inadvertently was not given the EFT.

Hypothesis 1. As predicted, females who identify with the father have significantly lower EFT scores than those who identify with the mother (Table 1). For males and both sexes combined, the difference is in the predicted direction but fails to reach the .05 significance level. It should be pointed out that contrary to expectations, more women than men identified with the father, and more men than women identified with the mother (Table 1). These characteristics of the *S*s' identifications are consistent with the interpretation that the males in this study tend to be relatively field dependent and the females tend to be field independent, and that these tendencies contributed to the minimal over-all sex differences in EFT performance that were reported above.

Hypothesis 2. Each sex group was divided into thirds, representing high, middle, and low AA scores. Hypothesis 2 was evaluated in two ways; first, by testing the significance of the difference between EFT scores of high and low AA *S*s, and second, by correlating AA and EFT scores for *all* *S*s in each sex group. It should be noted that the combined-sex groups of 20 high AA *S*s and 19 low AA *S*s in Table 1 are almost identical to the groupings that would have resulted had all 60 *S*s been placed in a single distribution and then divided them into high, middle and low AA groups. By this latter method, three males in the middle group would be in the high group and three females in the high group would be in the middle group. The combined-sex groups employed avoid the unequal distribution of males and females that would have resulted by this method. From Table 1 it is apparent that for both sexes *S*s high in AA had higher EFT scores than *S*s low in AA, although this difference is significant only in the case of males. When the sexes are combined, the difference is significant at the .04 level. Similar results for this hypothesis were obtained by the correlational analysis. For males, AA and EFT were significantly correlated ($r = .38$, $p = .05$), while for females the correlation was negligible ($r = .05$). For the combined sexes, the value of r was .28 ($p = .05$).

Hypothesis 3. This hypothesis proved difficult to test adequately because of the few *S*s who fell into the groups needed. Previous research (Bieri & Lobeck, 1959) has indicated that *S*s high in AA tend to identify more with the father in contrast to low AA *S*s who identify more with the mother. Therefore, fewer *S*s would have high AA scores and maternal identification or low AA scores and paternal identification. In spite of the few cases obtained, differences in the predicted direction were obtained (Table 1), although these differences were significant only when males and females were combined.

DISCUSSION

While the results were consistently in the predicted direction, the findings suggest that parental identification and acceptance of authority may operate differentially within each sex in relation to EFT performance. For females, parental identification seems more reliably associated with differences on the EFT than does high or low AA. This conclusion is in line with previous work with these same *S*s (Bieri & Lobeck, 1959) which suggests that AA is a less discriminating variable in women than in men, possibly because attitudes toward authority play a more important role in the male's sex-role development.

For men, AA may be a more discriminating variable in terms of EFT performance than is parental identification. Thus, while high AA males had significantly higher EFT scores than low AA males, the most striking contrast in EFT performance was between males with high and low AA scores who identified with the mother. The mean EFT scores for these two groups were 67.25 and 18.00, respectively, a difference significant below the .01 level. In contrast, males who identify with the father but who differ in terms of having high or low AA scores, have mean EFT scores that are not reliably different (25.17 and 28.00, respectively). Thus, in both males and females, the combination of high AA and mother identification produces the most field-dependent behavior. For males, low AA and mother identification produce the most field-independent behavior, while for females, low AA and father identification produce the most field-independent behavior.

The greater field independence observed in *S*s with cross-sex identification and low AA is perhaps consistent with the interpretations sometimes made that persons considered to have a more stable sexual identification may manifest some opposite-sex tendencies in their behavior. Thus, Witkin et al. (1954) found that field-independent *S*s had more opposite-sex features in their drawings of same-sex humans than did field-dependent persons. Similarly, Adorno et al. (1950) consider the low authoritarian to be more secure in his sexual identification and thus able to acknowledge more opposite-sex behavior than the high authoritarian.

These findings make it desirable to revise the original assumptions of the study concerning the nature of identification in relation to cognitive behavior. The results are generally consistent with the reasoning that father identifiers are more field independent and mother identifiers are more field dependent. However, the fact that low AA males who identified with the mother had significantly lower EFT scores than high AA males who identified with the mother suggests that maternal identification in males reflects a more passive, field-dependent orientation when combined with high acceptance of authority, but reflects a more active, field-independent orientation when combined with low acceptance of authority. Thus, not only do low AA males with maternal identification have lower EFT scores than high AA males with this identification, but in addition they are the only male group whose mean EFT score is significantly lower than the mean of the comparable female group (18.00 and 39.00, respectively; $p = .02$, two-tails).

It is apparent that due to the restricted nature of the samples of this study, and the small numbers of *S*s involved, the present findings are best construed as providing direction for further research. In particular, it should be stressed that the study bears on only one aspect of identification, perceived similarity between oneself and one's parents. The results suggest that additional work along the lines of this study may not only yield additional information on the relationship of identification to cognitive behavior, but also clarify the nature of the identification process itself.

SUMMARY AND CONCLUSIONS

Sex differences in cognitive behavior may be conceptualized in terms of personality variables related to within-sex differences, when it is reasonable to assume these variables are related to the cognitive behavior under study. Using measures of parental identification and acceptance of authority, it was predicted that within each sex, and for both sexes combined, the following differences in performance on the embedded figures test should be found: (*a*) *S*s who identify with the father should score lower on the EFT than mother identifiers; (*b*) *S*s low in AA should score lower on the EFT than *S*s who are high in AA; (*c*) *S*s with low AA scores and who identify with the father should score lower on the EFT than *S*s with high AA scores and who identify with the mother.

The results support Hypothesis 1 in the case of females only, although for males and for the sexes combined, the results are in the predicted direction. Hypothesis 2 is supported for males and for both sexes combined, and again the results for females are in the predicted direction. Hypothesis 3 is supported for sexes combined, although differences in the predicted direction were obtained for males and females separately.

These results suggest that parental identification may operate more simply in relation to differences in EFT performance for females, while attitudes toward acceptance of authority may operate more directly in relation to EFT performance in males. However, further analyses indicated that for males, *S*s with low AA and maternal identification had significantly lower scores than *S*s with high AA and maternal identification. Thus, maternal identification in males may reflect quite different cognitive orientations depending upon the level of attitudes toward authority. The relevance of this cross-sex identification in behavior was discussed in relation to these findings. Further research with these variables might add to our understanding of both the nature of personality dispositions affecting cognitive behavior and the nature of identification per se.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BALES, R. F., & COUCH, A. A factor analysis of values. Unpublished manuscript. *Lab. of Soc. Relat.*, Harvard Univer., 1956.

BIERI, J., BRADBURN, WENDY, M., & GALINSKY, M. D. Sex differences in perceptual behavior. *J. Pers.*, 1958, **26**, 1–12.

BIERI, J., LOBECK, ROBIN, & GALINSKY, M. D. A comparison of direct, indirect, and fantasy measures of identification. *J. abnorm. soc. Psychol.*, 1959, **58**, 253–258.

BIERI, J., & LOBECK, ROBIN. Acceptance of authority and parental identification. *J. Pers.*, 1959, **27**, 74–86.

JENSEN, A. R. Personality. In P. R. Farnsworth (Ed.), *Annu. Rev. Psychol.* Palo Alto, Calif.: Annual Reviews, 1958. Pp. 295–322.

MARLOWE, D. Some psychological correlates of field independence. *J. consult. Psychol.*, 1958, **22**, 334.

WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

(Received November 24, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE EFFECTS OF NEED-ORIENTED COMMUNICATIONS ON ATTITUDE CHANGE[1]

FRANCIS J. DI VESTA AND JACK C. MERWIN
Syracuse University

One of the processes underlying attitude change appears to consist of the internalization, or incorporation, of the ideational content of the attitude (Katz, Sarnoff, & McClintock, 1956). A theoretical framework based on this position (Woodruff & Di Vesta, 1948; Smith, 1949; Rosenberg, 1956) assumes that favorable attitudes toward an object (issues, activities, ideologies, and the like) result from perceptions (beliefs or accepted information) that the attitude object facilitates need satisfaction; and unfavorable attitudes result from perceptions that the attitude object blocks or hinders need satisfaction. Accordingly, it follows that the manipulation of need (or value) strength, of the perceived instrumental relationship of the object to the satisfaction of needs, or of both may be effective in producing a change in attitudes.

The present experiment differed from earlier investigations (Woodruff & Di Vesta, 1948; Carlson, 1956; Rosenberg, 1956; Peak & Morrison, 1958) based on this theory of attitude structure in the following ways: (*a*) change in attitude was attempted by increasing the *S*'s awareness of the instrumentality of the attitude object for attaining a specified need rather than an indirect orientation toward multiple needs or values; (*b*) both negative and positive instrumental orientations were provided, each in a separate communication; (*c*) in both communications a generally positive affective assertion toward the attitude object was maintained; (*d*) changes in instrumentality of the attitude object for satisfying the need engaged in the communication were compared to corresponding changes for needs not so engaged; (*e*) the total range of scores was involved in the analyses rather than only extreme scores; and (*f*) all measures were obtained by summated rating scales developed in accordance with the definitions emerging from the theory.

The specific hypotheses were as follows:

H1: If a communication about a given attitude object or event engages a given need, the extent to which *S*s perceive the attitude object to be instrumental for satisfying the need can be changed.

- 1(a) If the communication engages a need in such a way that the attitude object is viewed as favorable to the satisfaction of the need, the change in perceived instrumentality will be positive.
- 1(b) If the communication engages a need in such a way that the attitude object is viewed as unfavorable to the satisfaction of the need, the change in perceived instrumentality will be negative.

H2: The change in perceived instrumentality of the attitude object for satisfying the need engaged by the communication will be greater than the corresponding change for any other need.

- 2(a) If the need is engaged by the communication in a manner favorable to the need, there will be a greater positive change in the perceived instrumentality for that need than for any other need.
- 2(b) If the need is engaged by the communication in a manner unfavorable to the need, there will be a greater negative change in the perceived instrumentality for that need than for any other need.

H3: The correlation between change in attitude and change in affect induction (a

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 41(657)-73 monitored by the Office for Social Science Programs, Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, and disposal in whole and in part by or for the United States Government. This is not an official publication under the contract. Views or opinions expressed or implied herein are not to be construed as necessarily reflecting the views or indorsement of the Department of the Air Force or of the Air Research and Development Command.

multiplicative combination of need strength and perceived instrumentality) will be greater for the need engaged by the communication than for any other need.

METHOD

Subjects

The *S*s in the main experiment were 218 freshmen enrolled in the College of Liberal Arts at Syracuse University. An additional 300 *S*s from the same population, 87 *S*s from the School of Education, and 155 high school upperclassmen were used in the various studies of reliability and validity of the scales. Eighty-six *S*s enrolled in a state teacher's college were used in a pilot study.

Procedure

The attitude object for the experiment was that of "teaching as a career." *S*s in the main experiment were volunteers who knew nothing about the purpose of the study. They were part of a larger group who had taken the perceived instrumentality scales, the attitude scale, and a needs scale, together with the testing program administered to all freshmen. Upon volunteering they were alternately assigned to one of two experimental groups or the control group. The experiment was carried out in seven sessions distributed over a three-day period. During each session there were three groups of *S*s; one group received a positive communication, one a negative communication, and one a control communication. Since three *E*s were used, the communications were rotated after each session to distribute *E* effects randomly.

At the beginning of the experimental session *S*s were informed that their help was solicited to evaluate an educational radio program, and that they would be asked for reactions to questionnaires both before and after the program was played. *S*s then responded to instruments measuring needs and attitudes. These papers were collected and the program was played. At the conclusion of the program *S*s completed the perceived instrumentality and attitude scales. A detailed description of theoretical orientation, design, and measures is available elsewhere (Di Vesta & Merwin, 1958).

Measures

Attitude scales. Our theoretical orientation distinguishes between attitude and perceived instrumentality (PI). Each of the items written for the attitude scale was screened for maximum affective and minimum cognitive content. An attitude was thus defined as a relatively stable tendency to respond with positive or negative affect to a specific referent. An example of the attitude items is: "I am sure I would enjoy teaching." *S* was instructed to tell how he personally felt about teaching as a career for himself by checking one of the six categories ranging from strongly agree to strongly disagree for each of 11 statements. Statements expressing favorable opinions and statements expressing unfavorable opinions were included.

TABLE 1

INTERCORRELATIONS OF PERCEIVED INSTRUMENTALITY AND ATTITUDE MEASURES

($N=155$)

PI	Perceived Instrumentality			Attitude
	Affiliation	Dominance	Exhibition	
Achievement	.44	.44	.46	.24
Affiliation		.22	.39	.29
Dominance			.44	.14
Exhibition				.26

The split-half reliability coefficient (300 college freshmen) corrected by the Spearman-Brown formula was .71. The coefficient of stability of the scale was .79 for 218 freshmen over a four-month period.

Perceived instrumentality scales. The measure of this aspect of cognitive structure was obtained by the use of an instrument consisting of four separate summated rating scales. These were measures of the extent to which the *S* believed that career teaching facilitated or hindered the satisfaction of the need for achievement (PI Ach), of the need for affiliation (PI Aff), of the need for exhibition (PI Exh), and of the need for dominance (PI Dom). Each scale was composed of statements about which *S* was instructed to indicate his agreement or disagreement on a six-point scale from strongly agree to strongly disagree. An illustration of a typical item is as follows: "Teachers have more opportunity to be creative than do people in most other occupations." Items were structured to obtain the *S*'s *evaluative* behavior rather than his preferential behavior as for attitudes.

The scale concerning n Ach was used as a measure of the effectiveness of the change procedure since the communications engaged that need. The other measures of perceived instrumentality (PI Dom, PI Exh, PI Aff) were used as controls.

Initially, a 64-item instrument was developed with the items for the four scales intermixed at random. Each of the four scales included some statements indicating that career teaching had a high degree of instrumentality for satisfying the need and other statements indicating that it had a low degree of instrumentality. The instrument was administered to 155 senior high school students together with the attitude scale. The intercorrelations among the four scales and the correlation of each with the attitude scale are presented in Table 1. These coefficients indicate the relative independence of these measures.

Evidence for the validity of the scale for measuring PI Ach was obtained in the administration of the 64-item scale to 87 college students majoring in education. The hypothesis was that persons having selected teaching as a career have done so, in part, on the basis of expected satisfactions of n Ach. The mean item score on the PI Ach scale was predicted to be higher for the education majors than for the high school group. The difference in the means ($t = 4.56$) was in the predicted direction and significant ($p < .001$).

The final form of the perceived instrumentality instrument consisted of 56 of the original 64 items. A coefficient of stability was calculated for each of the four 14-item subscales. The coefficients, based on data obtained from 73 college freshmen retested after a four-month period, were: .59 for PI Ach, .55 for PI Aff, .70 for PI Dom, and .65 for PI Exh.

Need scales. Two 40-item Likert scales for measuring the four needs were developed, patterned after statements in the *Activities Index* (Stern, 1956) and the *Edwards Personal Preference Schedule* (Edwards, 1957). One of these scales consisted of statements describing activities a person would perform in social, intellectual, and/or career situations. Using a six-point scale, *S* indicated the extent to which he would *seek* or *avoid* situations involving each activity. Each of the four subscales was composed of 10 items, and the 40 items were intermixed at random throughout the instrument. On the second instrument, 40 additional statements of activities were used. The *S* used a six-point scale to indicate, for each of these statements, the extent to which he *liked* or *disliked* the activity involved.

The equivalence of the subscales was determined by relating the corresponding scores from the two newly developed instruments. The coefficients of correlation, based on 218 college freshmen were: .62 for n Ach, .57 for n Aff, .70 for n Dom, and .68 for n Exh.

Communications

Three communications were used. One provided information that teaching had favorable consequences for satisfying the need for achievement; the second that teaching had unfavorable consequences for satisfying it; and the third was a control communication on a related issue but did not engage needs in any direct manner. In both of the experimental communications a favorable assertion toward teaching was maintained, i.e., "teaching is a good career because . . . ."

The experimental communications were prepared by the *E*s and recorded on tape by the Department of Radio and Television of Syracuse University. They were the same with respect to: (*a*) the authoritative moderator, (*b*) a panel of senior division college students, (*c*) primary sources used as the bases for comments, (*d*) the number of points covered, (*e*) the enthusiasm of speakers, and (*f*) length. All programs were introduced by an announcer of the local ratio station.

Program I. In this presentation several sources were quoted by members of the panel to provide information that teaching has positive instrumentality for satisfying n Ach. Following the statement given by each member was a short summary by the moderator. Typical statements were as follows:

> The counsellors also pointed out that teaching is one job where you can get concrete evidence of your accomplishments.
>
> . . . the teacher is continually making advancement into new frontiers.
>
> . . . teaching is an occupation involving personal accomplishment, growth, and contribution to the community.

Throughout the discussion there was continual give-and-take between the panel members (college students) and the moderator (a college professor). The running time was 14 min.

Program II. This program involved the same participants and format as Program I. The questions asked by the moderator were the same as in the first program. The orientation of all information presented was that teaching lacks instrumentality for satisfying n Ach. For example, two of the statements were:

> (*a*) (In one survey) all teachers mentioned that one of the main advantages of the teaching profession, unlike other professions, is the emphasis placed on the teacher doing his own job rather than being continually in competition with other teachers; and,
>
> (*b*) The teacher has many vacation periods during which she can rest and forget about her job—unlike the scientist who continually strives to achieve new discoveries, or the industrial engineer who may spend his vacation trying to improve a product, or even the politician who during his vacation tries to achieve a bigger following.

Again it should be noted that this communication was directed toward advantages of teaching. Like the first program, this program was aimed at "why teaching is a good career." The two programs differed only in the instrumental orientation of information presented. The running time for this program was 14 min.

Program III. A control group listened to a tape recording on the topic "Going to College." This was a question and answer program that dealt with such topics as large universities vs. small colleges, costs of going to college, obtaining college catalogs, and the like. The running time was 14 min.

Pilot Study of Programs

A pilot study of the programs was conducted with 86 freshmen of a state teacher's college. The *S*s were assigned at random to one of three groups. One of the groups listened to the positive communication and a second to the negative communication. Immediately following this activity the two experimental groups responded to the PI Ach and attitude scales. The third group responded to the same scales but did not listen to a program. No pretests were used in the pilot study.

The mean PI Ach and attitude scores for the three groups are presented in Table 2. An analysis of variance of PI Ach scores yielded an F of 10.96 ($p < .001$). The t values for the differences between means of the positive and negative groups ($t = 4.05$) and between the means of the control and negative groups ($t = 3.68$) were both significant ($p < .001$). The difference between means of the positive and control groups ($t = .80$) was not significant ($p > .05$). In view of these findings and inasmuch as the means were in the pre-

TABLE 2

MEANS OF PI ACH AND ATTITUDE SCORES ON PILOT STUDY

Experimental Treatment	*N*	PI Ach	Attitude
Positive communication	29	40.83	23.75
No communication	28	42.43	23.95
Negative communication	29	50.39	23.21

Note.—The lower the mean the more favorable the PI or attitude.

dicted order the programs were considered adequate for altering PI Ach.

An analysis of variance of attitude scores yielded an insignificant *F* of .75. This finding can probably be attributed to the fact that the *S*s, having selected teaching as a career and enrolled in a teacher's college, were committed to a highly favorable attitude toward teaching.

RESULTS AND DISCUSSION

The initial homogeneity of the three groups used in the main experiment is demonstrated by the *F* values obtained through an analysis of variance of the scores from the pretest measures. The *F* value for the PI Ach measure was significant ($.05 > p > .025$), but did not cause undue complications since the obtained order of the means (positive-control-negative) was opposite to that predicted to result from the experimental procedures. A significant *F* value ($.05 > p > .025$) for the PI Dom measure was not attributable to differences in means alone since the chi square value obtained from the Bartlett test for homogeneity of variance was also significant ($p < .05$). The *F* values for the other pretest measures used in the analyses were not significant. On the basis of the evidence obtained, the groups were judged to be sufficiently homogeneous for the experiment.

TABLE 3

COMPARISON OF GROUPS ON POSTTEST MEASURES

Measure	Experimental Group					
	Positive ($N = 76$)		Control ($N = 73$)		Negative ($N = 69$)	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Perceived Instrumentality						
Achievement	45.39	9.10	47.99	7.70	52.45	9.87
Affiliation	45.71	7.44	47.03	6.25	47.93	8.26
Dominance	41.87	7.30	43.45	8.36	44.16	7.58
Exhibition	44.12	6.40	45.25	7.23	47.32	6.81
Attitude	34.97	10.79	36.30	8.86	38.20	11.49

TABLE 4

MEANS AND *t* VALUES[a] OF CHANGES IN PERCEIVED INSTRUMENTALITY SCORES

Measure	Experimental Group					
	Positive		Control		Negative	
	$\bar{X}$	*t*	$\bar{X}$	*t*	$\bar{X}$	*t*
Perceived instrumentality						
Achievement	3.33	3.67***	0.62	0.67	−7.10	−5.87***
Affiliation	2.25	2.56*	0.10	0.12	−0.91	−0.92
Dominance	1.37	1.23	1.11	1.44	−2.73	−2.12*
Exhibition	2.01	2.81**	1.19	1.65	−3.45	−4.45***

[a] Mean changes tested against a hypothesized population mean of zero ($H: \mu = 0$).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

The means and standard deviations of scores on the posttests are presented in Table 3. The mean changes for the three groups on the PI scales are given in Table 4. A plus change in PI indicates a move toward the position advocated in the positive communication and a minus change indicates a move toward the position advocated in the negative communication. "Boomerang" effects occurred with some *S*s in both groups.

Each of the mean changes was evaluated by a *t* test, against the statistical hypothesis that $\mu = 0$, in accordance with experimental hypotheses H:1(a) and H:1(b). As may be seen in Table 4 the changes for both the positive and negative groups were significant ($p < .001$). The mean change of the control group was not significant ($p > .05$). These results permit rejection of the null hypothesis and clearly support experimental hypotheses H:1, H:1(a), and H:1(b).

The mean differences between change in PI Ach and change in each of the remaining three PI scores were all in the predicted direction for the positive group. None was significant however. While these data indicate trends, they do not provide sufficient evidence to support experimental Hypothesis 2(a).

For the negative group, the mean differences between change in PI Ach and the change in each of the other three PI scores were all significant ($p < .01$) and in the predicted direction. Thus, the data support experimental Hypothesis 2(b) and permit rejection of the null hypothesis of no difference in the changes.

Further support of the general experimental hypothesis (H:2) is the lack of significant differences in the changes for the control group.

In general, the effects of the communications on the PI Ach measures were as hypothesized. Nevertheless, a generalization effect is to be noted in which the positive communication

TABLE 5

ANALYSIS OF VARIANCE OF ATTITUDE CHANGE

Source	*df*	*MS*	*F*
Between groups	2	112.73	7.05*
Within groups	215	15.99	
Total	217		

* $p < .001$.

TABLE 6

MEANS AND *t* VALUES OF CHANGES IN ATTITUDE

Experimental Group	*N*	$\overline{X}$	t^a
Positive	76	1.05	2.15*
Control	73	−0.81	−1.83
Negative	69	−1.29	−2.80**

[a] $H: \mu = 0$.

* $p < .05$.

** $p < .01$.

produced a generally positive change, and the negative communication produced a generally negative change in PI.

The null hypothesis that there are no significant differences in attitude change among the three groups over the experimental period was tested by an analysis of variance summarized in Table 5. The obtained *F* value of 7.05 was significant ($p < .001$). Since the chi square from the Bartlett test of homogeneity of variance was not significant, the null hypothesis was rejected. The mean and *t* values of the attitude changes for the experimental groups are presented in Table 6. The mean changes in attitude were in the direction of the instrumental positions advocated in the negative and positive communications. Since positive affect was asserted in both communications, the change in attitude is attributed to the manipulation of perceived instrumentality. These effects are further substantiated by the fact that the mean change in the attitudes of the control group was not significant ($p >$ .05). The results indicate that attitudes can be modified by a manipulation of the instrumental orientation of the attitude object for attaining or hindering the satisfaction of a single need.

A measure of the change in affect induction (AI) for each need was obtained by multiplying the need score with the corresponding PI change for each individual. For example, the score for n Ach was multiplied by the change in PI Ach; n Dom was multiplied by the change in PI Dom; and so on. In addition, a composite AI change score was obtained by an algebraic summation of the four simple AI change scores. Each AI change score was correlated with the attitude change for each treatment group and for the three groups pooled. The resulting coefficients are presented in Table 7.

TABLE 7

CORRELATIONS BETWEEN CHANGE IN ATTITUDE AND CHANGE IN AFFECT INDUCTION

Affect Induction	Experimental Group			
	Positive ($N = 76$)	Control ($N = 73$)	Negative ($N = 69$)	Pooled ($N = 218$)
Achievement	.26*	.07	.22	.26**
Affiliation	−.09	−.20	−.04	−.10
Dominance	.13	−.06	.00	.07
Exhibition	.14	−.09	.10	.11
Composite	.17	−.12	.11	.13

* $p < .05$.

** $p < .001$.

Only the correlations between change in affect induction for n Ach and attitude change for the positive and pooled groups were significant. While these results would appear to provide partial support for Hypothesis 3 the relationships are too low to be considered conclusive. While the hypothesis was not advanced in the present study, it would be consistent with the theory for the composite change in affect induction to be more highly related to attitude change than would changes in affect induction based on single needs. This expectation was not supported by the data and leads to a conclusion similar to that drawn by Rosenberg (1956, p. 371) for his study that "The data leave it an open question as to whether . . . 'perceived instrumentality' actually controls more variance in attitudinal affect than does 'value [need] importance.' "

SUMMARY

Several hypotheses concerning attitude change and derived from a cognitive theory of attitude structure were tested experimentally. The change procedure consisted of three tape recorded communications, each making positive affective assertions about teaching but

maintaining respective positions that the attitude object facilitated or hindered satisfaction of the need for achievement, or making assertions irrelevant to both teaching and n Ach. Each of the two experimental groups and a control group consisted of approximately 70 subjects. Changes in beliefs about the potentiality of the attitude object for satisfying the need for achievement corresponded to the direction of the instrumental position of the respective communications. Corresponding changes in attitude were demonstrated. An index based upon need strength and change in perceived instrumentality for the need engaged in the communication was found to be only slightly related to change in attitude for the pooled groups. Similar indexes based on other needs and over all needs measured were found not to be related to shifts in attitude.

REFERENCES

CARLSON, E. R. Attitude change and attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.

DI VESTA, F. J., & MERWIN, J. C. The effect of need-oriented communications on attitude structure and change. USAF WADC tech. Rep., 1958, No. 58-93 (ASTIA Doc. No. AD-151-044).

EDWARDS, A. L. *Edwards Personal Preference Schedule.* New York: Psychological Corp., 1953.

KATZ, D., SARNOFF, I., & MCCLINTOCK, C. Ego-defense and attitude change. *Hum. Relat.*, 1956, **9**, 27–45.

PEAK, HELEN, & MORRISON, H. W. The acceptance of information into attitude structure. *J. abnorm. soc. Psychol.*, 1958, **57**, 127–135.

ROSENBERG, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367–372.

SMITH, M. B. Personal values as determinants of a political attitude. *J. Psychol.*, 1949, **28**, 477–486.

STERN, G. G. *Activities Index.* Syracuse: Psychological Res. Center, Syracuse Univer., 1956.

WOODRUFF, A. D., & DI VESTA, F. J. The relationship between values, concepts, and attitudes. *Educ. psychol. Measmt.*, 1948, **8**, 645–660.

(Received December 19, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

INVOLVEMENT AND COMMUNICATION DISCREPANCY AS DETERMINANTS OF OPINION CONFORMITY[1]

PHILIP G. ZIMBARDO
Yale University

A MAJOR problem of theoretical and practical significance in the area of attitude change concerns the placement of a persuasive communication relative to the attitude position of the person to be influenced. The bulk of experimental evidence points to a positive relationship between attitude change and degree of change advocated (Fisher, Rubinstein, & Freeman, 1956; Fisher & Lubin, 1958; Goldberg, 1954; Harvey, Kelley, & Shapiro, 1957; Hovland & Pritzker, 1957; Hovland, Harvey, & Sherif, 1957). Thus, the more discrepant a communication is from the position of the recipient the greater the attitude change in the advocated direction. However, it has been suggested by Hovland et al. (1957) that this generalization may be limited to communications (*a*) within the subjects' latitude of acceptance, (*b*) about noninvolving issues, (*c*) by credible, respected communicators.

This view stems from an extrapolation by Hovland and Sherif of results from a psychophysical weight-lifting study (Sherif, Taub, & Hovland, 1958) to the situation of communication and attitude change. They feel that discrepant communications within the latitude of rejection act like anchors *beyond* the range of a series of stimuli and should produce contrast or negative change, while communications within the latitude of acceptance act like anchors *within* a series of stimuli and should produce assimilation or positive change.

A latitude of acceptance is comprised of the position on an attitude continuum that an individual "most accepts" plus those he also accepts. The latitude of rejection includes the attitude position to which the individual "most objects" plus those to which he also objects. According to Hovland and Sherif, then, it is the placement of a communication relative to the individual's latitudes of acceptance and rejection that determines the efficacy of the communication.

A more general theoretical framework in which to explore the motivational bases of opinion change in situations of social influence has recently been advanced in Festinger's (1957) theory of cognitive dissonance. A number of derivations can be made from the theory which predict the relationship between opinion change and the variables that Hovland and his co-workers have isolated empirically.

Dissonance theory assumes a basic tendency toward consistency of cognitions about oneself and about the environment. When two or more cognitive elements are psychologically inconsistent, dissonance is created. Dissonance is defined as psychological tension having drive characteristics. Thus, the existence of dissonance is accompanied by psychological discomfort and when dissonance arises, attempts are made to reduce it. A specific application of the theory to the problem under consideration in the present paper states that dissonance is created when an individual knows that another person—a communicator—holds an opinion contrary to his own (Festinger, 1957, p. 178). The magnitude of the dissonance created by the expression of a contrary opinion will increase with increases in (*a*) the credibility and attractiveness of the communicator, (*b*) the extent of the discrepancy in opinion, and (*c*) the importance of, or involvement with, the cognitive elements of the situation.

The strength of the pressure to reduce dissonance is a direct function of the dissonance created. In a social influence situation there are a number of potential channels of dissonance reduction, such as changing one's own opinion, changing the opinion of the communicator, making the communicator noncomparable to oneself, seeking further support for one's position, dissociating the source from the content of the communication, and distorting the meaning of the communication.

---

[1] This research is part of a dissertation presented to the faculty of the Graduate School of Yale University in candidacy for the degree of doctor of philosophy.

The author wishes to thank A. R. Cohen for the guidance he provided during every phase of this research.

Funds from the Yale Communication and Attitude Change Project, directed by C. I. Hovland, helped to support this study.

The author is indebted to Sister Thomas Albert, Dean of Albertus Magnus College, for her cooperation throughout the study.

One of the major methods of reducing dissonance, then, is conformity or opinion change in the advocated direction. Thus, if the other avenues of dissonance reduction are restricted, opinion change should be a function of variables that intensify dissonance. We have seen that two such variables are opinion discrepancy and importance of, or involvement with, the cognitive elements of the situation.

While discrepancy and involvement play key roles in the theory, nevertheless they remain rather vague and insufficiently specified. For the present study, therefore, discrepancy was operationally defined in Hovland and Sherif's terms as latitudes of acceptance and rejection.

The variable of involvement, which has been used by most investigators to mean concern with a given *issue* because it is intrinsically involving and is related to the individual's needs and values, was used in a different sense in the present study. This naturalistic kind of issue-involvement often confounds a number of factors, such as initial differences in commitment, social support, information, etc.

Specification of involvement in terms of the relationship of one's opinion to the achievement of a desired goal should help to limit ambiguity of interpretation. This response-involvement may be viewed as the individual's concern with the consequences of his *response* or with the instrumental meaning of his opinion. The involvement may then be differentially imposed on a previously neutral opinion area, thereby controlling the contaminating factors usually associated with the selection of different levels of issue-involvement.

Thus, the following predictions can be generated by dissonance theory for the situation in which an individual receives a persuasve communication from a "positive" communicator (response alternatives other than opinion change having been experimentally minimized): (*a*) dissonance, and hence opinion conformity, will be greater for those individuals who are highly involved than for those whose involvement is low; (*b*) dissonance, and hence opinion conformity, will be greater when the communication is widely discrepant and is directed at an individual's latitude of rejection than when it is slightly discrepant and within his latitude of acceptance; (*c*) maximum dissonance, and consequent opinion conformity, will occur when the individual is highly involved (Hi-I) and the communication is within his latitude of rejection (Lat-Rej); (*d*) minimum dissonance, and consequent opinion conformity, will occur when the individual's involvement is low (Lo-I) and the communication is within his latitude of acceptance (Lat-Acc).

In the fourfold table generated by these two variables, there should be an intermediate amount of dissonance and consequent opinion change in the cells in which the main effects are mixed, i.e., high involvement with a latitude of acceptance communication, and low involvement with a latitude of rejection communication. The theory cannot predict which of the two will be more effective, since that depends on the relative efficacy of the experimental manipulation of involvement and discrepancy.

METHOD

In order to test these predictions a before-after design was used in which 80 unpaid college girls privately gave their opinions about the locus of blame in a case study of a juvenile delinquent. After their involvement was experimentally manipulated, they were exposed to the alleged opinion about the case of a girl friend with whom they were being tested. Finally, they gave their opinions of the case study again.

Procedure

Girls volunteered with a friend for a study on how friends diagnose and evaluate social issues. The written responses of each subject (*S*) of the two pairs tested simultaneously were kept confidential by means of short partitions.

A cooperative task designed to introduce the *Ss* to the situation and to highlight the cohesiveness of each pair was given first. This task required the *Ss* to list as many social problems as they could in 60 sec., each *S* alternately listing one problem. Regardless of how well they had done, they were told that their performance was good and was better than that of comparable *S*s who were not friends. Following this task, no communication was permitted between the *S*s at any time.

The *S*s then filled out a general background questionnaire which included checks on the strength of their friendship. At this time, *E* stressed the confidential nature of their responses. They were told that because of time limitations only one social problem was selected for them to work on in more detail. The problem of juvenile delinquency was selected and related to it were two tasks. The first task involved reading a short, fictitious case history of a juvenile delinquent—"the Johnny Rodriguez case." *S* was then required to indicate her opinion concerning the locus of blame for the youth's crime. This opinion was the one *E* attempted to change.

The second task, which involved judgments of delinquents from among a set of photographs, was included primarily to increase the credibility of the friend communicator.

Each *S* was informed that while she had judged only five of eight photos correctly and was rated "good," her friend had judged all of the photos correctly and had been rated "expert." The *S*s had been told that this task measured their perceptual and judgmental ability. By having the friend appear an expert on the photo task, the impact of her *opinion* about the case study could be maximized without resorting to additional information about the case.

Involvement Manipulation

Low (*Lo-I*). Half of the *S*s were told that although they should read the case study carefully, they should not expect too much of it because it was too short, limited, and not representative of juvenile delinquency in general. Therefore, it was impossible to learn anything from their reactions to the case. They were told that *E* had hoped to learn something about their basic values and personalities by means of their case study opinions, but results with other girls from the school indicated that this attempt had proven unsuccessful. However, in order to keep everything in the experiment in the same sequence, it was necessary that they read the case and give their opinions before judging the photos.

High (*Hi-I*). *E* told the other half of the *S*s that the case study they were to read was selected because, although short, it presented a wide range of problems implicitly or explicitly, and the way they responded to this typical, representative case would be a good indicator of their basic social values, their personalities, and their outlook on important life problems.

Both groups were told that the photo task would provide information of a different kind, about an ability rather than about their opinions. After reading the case material, the *S*s made the following ratings on a scale composed of nine statements (see below) which varied in the degree to which the blame for Johnny's crime was personal or environmental. They: (*a*) selected the one statement which was most acceptable (MA), i.e., which came closest to representing their own point of view; (*b*) selected all others they felt were also acceptable (A); (*c*) selected the one statement which was most objectionable (MO); (*d*) selected all other statements which were also objectionable (O); and, finally, (*e*) indicated all statements they were unsure of, or neither accepted nor rejected.

After the opinion rating sheets were collected, the *S*s indicated how *certain* they were of their most acceptable rating.

Opinions about Johnny

1. There is no question that the blame for Johnny's crime must be placed *entirely* upon his shoulders.
2. The *greatest part* of the blame for Johnny's crime rests on his shoulders.
3. While some of the blame for Johnny's crime is due to his background, he is still responsible for *most* of it.
4. Although it is hard to decide, it is probable that *slightly more* of the blame for Johnny's crime rests on his shoulders.
5. It is difficult to decide whether Johnny himself or his background is to blame for his crime.
6. Although it is hard to decide, it is probable that *slightly more* of the blame for Johnny's crime rests on his background.
7. While some of the blame for Johnny's crime is due to Johnny himself, his background is still responsible for *most* of it.
8. The *greatest part* of the blame for Johnny's crime rests on his background.
9. There is no question that the blame for Johnny's crime must be placed *entirely* upon his background.

Communication of Friend's Opinion

The *S*s were told that since they came to the experiment in pairs of friends they would probably like to know not only how they were doing but also how their friend was doing. Thus, each pair would receive information about how they were doing, but would not have any information about the other pair of *S*s being run simultaneously.

After receiving the false photo judgment scores, every *S* was informed (on a written form) that her friend was "absolutely certain" of her most acceptable rating of the case. Half of the *S*s were informed (on a rating sheet) that their friend's MA rating was slightly discrepant from their own MA rating and was at a position which they had previously indicated as being "acceptable" (Lat-Acc). The other half of the *S*s were informed that their friend's MA rating was widely discrepant from their own and at a position they had previously indicated as being "objectionable" (Lat-Rej). Conveying the information on a rating sheet similar to that originally filled out by each *S*, and simply marking the friend's alleged opinion MA, served to make salient the opinion discrepancy and reduced the possibility of misinterpretation of the communication.

In the latitude of acceptance condition the statement selected as the friend's MA was generally one scale position above or below *S*'s MA at a position *S* had initially marked "acceptable." An attempt was made to avoid attributing a neutral position to the friend. For the latitude of rejection condition the statement selected as the friend's MA was, on the average, three scale units above or below *S*'s MA, at a position *S* had initially marked "objectionable". The friend's MA was generally placed on that side of the S's MA which offered the most potential scale movement. The Ss were randomly assigned to the discrepancy condition, while the involvement condition was alternated from session to session; all Ss within each testing session were given the same involvement induction.

In order to reduce the possibility that Ss would think that they could influence their friend to change *after* the experimental session, they were told that a follow-up questionnaire would be administered some time before the end of the term. Consequently, it was necessary for them not to talk to anyone about their reactions to any part of the study, especially not with their friend.

Postcommunication Ratings

Several minutes after having collected the false information, *E* said that the *S*s were to make a fresh evaluation of the case and of the photos. This was to be done because their first opinions and judgments were immediate, almost forced responses, and *E* wanted them to make new evaluations after they had considered the problem and materials in more detail. They were told that it did not matter whether or not they changed their opinions as long as they tried to make new, independent evaluations and did not try to duplicate what they had done merely to be consistent.

The *S*s then gave their opinions of the case study again, without rereading it. Checks were then made on *S*'s certainty of opinion, perceived discrepancy of friend's opinion, and degree of involvement in each of the tasks. The partitions were removed and *E* conducted an open-ended interview to inquire into the *S*s' reactions to the study. Finally, they were thanked for their cooperation and the true purpose and design of the study were revealed. They were asked (on their school "honor") not to discuss the study with any students who had not yet participated; and there is no reason to believe that the deceptions were not effective throughout the entire course of the study.

RESULTS

Evidence of the Effectiveness of the Experimental Conditions and Manipulations

Friendship. One of the basic assumptions of the study was that each pair of *S*s was very friendly. The friendship, along with the alleged expertness of the communicator, served to make the communicator highly positive. The data clearly confirm this assumption. The girls in each pair had known each other for an average of 30 months. In addition, they gave a mean rating of 1.17 on the friend "contact" scale (1.00 = "see and talk several times a day"); and made a mean rating of 1.38 on the friend "liking" scale (1.00 = "like very much"). These results are uniform throughout all experimental conditions. The fact that the *S*s publicly announced their friendship on the sign-up sheet on the bulletin board also supports this assumption.

Perception of opinion discrepancy. The *S*s in the Lat-Rej condition perceived the communication as being more discrepant than did the *S*s in the Lat-Acc condition, a difference significant beyond the .001 level. The differences in perceived discrepancy closely paralleled the actual, experimentally created differences in opinion.

Involvement in case study opinions. The *S*s in the Hi-I condition were significantly more concerned with their case study opinions than were the *S*s in the Lo-I condition. They also felt it was more important to make a sensitive judgment of the case (both p values $< .001$). No significant differences in degree of involvement were noted for the two latitude conditions.

The third check on the manipulation of involvement is a more sensitive one suggested by Hovland, Harvey, and Sherif (1957). Individuals highly involved in a given issue will have a raised threshold of acceptance of opinion statements about the issue, while less involved *S*s will have a lowered threshold of acceptance. The present results are in line with those of the above authors: highly involved *S*s reject a significantly greater number of statements than do less involved *S*s. In sum, these data indicate that the *S*s in the Hi-I condition were more involved with their case study opinions than were the Lo-I *S*s.

Involvement in photo judgments. Since the instructions regarding the photos were identical for all *S*s, it was expected that there would not be a significant difference between the two case study involvement conditions with regard to their photo involvement. The data support this assumption. One may conclude that the photo task was a method of raising the expertness of the communicator uniformly in all conditions without differentially affecting involvement with the case study opinions.

Certainty of opinion. Though the certainty measure was used in order to help anchor the communicator's position, it also provides a check on the *S*'s confidence in his own position. Since there were no significant differences among groups in mean certainty either before or after the communication we cannot attribute differences in opinion change to differences in certainty.

Effects of Discrepancy and Involvement on Opinion Change

In addition to use of traditional measures of opinion change, such as the scale distance changed and the percentage of *S*s changing in each direction, the scaling procedure used in the present study permits an analysis of opinion change into six components. Briefly, these positive opinion changes consist in (*a*) moving one's "most acceptable" rating in the advocated direction (*b*) moving one's "most objectionable" rating away from the advocated direction, (*c*) adding one (or more "acceptable" ratings in the direction of the communicator's position, (*d*) removing one (or more) "objectionable" ratings from positions near the communicator's position, (*e*) removing one (or more) "A" ratings from positions behind one's own MA rating, i.e., at a distance from the communicator's position, and, finally, (*f*) adding one (or more) "O" ratings to positions behind one's own MA.

It was arbitrarily decided to combine these components into a *composite change index* by assigning a weight of +1 to each type of positive change, −1 to each negative change and zero if no change occurred, and then summing algebraically. Thus, each individual could receive a composite change score varying from +6 (maximum positive change) to −6 (maximum negative change).

Composite change index. Figure 1 presents mean total change score as a function of position of the communication and strength of involvement. It is clear from the figure that (*a*) all groups showed positive change and (*b*) that opinion change increased both with increased discrepancy and with strength of involvement. The Hi-I, Lat-Rej group, in which the experimental effects were maximized, changed most, while the Lo-I, Lat-Acc group, in which the effects were minimized, changed least. The other two "mixed" groups changed an amount intermediate between these two extremes.

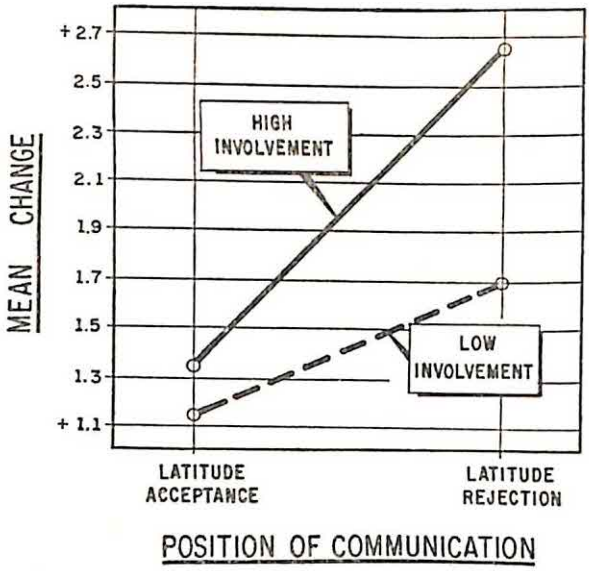


FIG. 1. Mean composite opinion change as a function of degree of involvement and extent of communication discrepancy. (Note.—Lat-Rej vs. Lat-Acc, $F = 7.74$, $p < .01$, $df = 1.76$. Hi-I vs. Lo-I, $F = 2.27$, $p = .10$, $df = 1.76$.)

An analysis of variance reveals that the difference between the two discrepancy conditions is reliable beyond the .01 level, and the difference between the two involvement conditions approaches significance at the .10 level. The Hi-I, Lat-Rej group differs significantly from each of the other groups.[2] The interaction between the main effects does not approach statistical significance.

Scale distance change in most acceptable rating. Table 1 presents an analysis of the data by number of scale units of net change in MA rating for each of the groups, and offers results similar to those obtained with the composite change index. There was significantly greater scale distance change in the advocated direction by *S*s who were highly involved than by *S*s whose involvement was low. Furthermore, significantly more change was produced by widely discrepant, latitude of rejection communications than slightly discrepant latitude of acceptance communications. Again, the relative ranking of the four cells was in accord with the predicted relationships among the groups.

Examination of the data of mean units of positive change in MA for only those *S*s who

---

[2] Hi-I, Lat-Rej vs.: Lo-I, Lat-Acc, $t = 3.19$, $p < .01$; Hi-I, Lat-Acc, $t = 2.77$, $p < .01$; Lo-I, Lat-Rej, $t = 2.02$, $p < .05$, $df = 38$. All p values are for two-tailed tests of significance.

TABLE 1

Mean Units of Net Change in "Most Acceptable" Rating

		Position Advocated		
INVOLVEMENT		LAT-ACC	LAT-REJ	
	Hi-I	.35	.95	.65
	Lo-I	.05	.30	.18
		.20	.62	

Note.—LAT-REJ vs. LAT-ACC, $F = 5.25$, $p < .05$, $df = 1.76$; Hi-I vs. Lo-I, $F = 4.19$, $p < .05$, $df = 1.76$.

actually changed their MA in a positive direction revealed trends identical with those presented for the entire groups.

Percentage of Ss changing most acceptable opinion. The results of net percentage of opinion change are in close agreement with the previous findings, e.g., the maximum dissonance group had 50% net change, while the minimum dissonance group had only 15% net change.

Equation of groups on opportunity for scale movement. Of course, the possibility must be entertained that the results for the discrepancy conditions could be explained by the methodological artifact that there was more scale distance available to change for the high discrepancy (Lat-Rej) groups than the low discrepancy (Lat-Acc) groups. Reanalysis of all the data after identically equating the groups on the degree of scale movement possible indicates that the differences obtained are accentuated rather than attenuated. For example, the difference between the discrepancy conditions on the composite change measure is significant beyond the .001 level, while the difference between the involvement condition is significant beyond the .08 level.

Relation of initial latitude size to opinion change. An unexpected finding revealed that those *S*s who conformed initially accepted a wider range of statements and rejected a narrower range than did the nonconformers, i.e., they had larger latitudes of acceptance and smaller latitudes of rejection. This measure of "threshold of acceptance," which significantly distinguished between positive changers and those changing negatively or not at all ($p < .01$), may thus prove to be a useful indicator of susceptibility to persuasive communications.

Discussion

The results of this experiment appear consistent with several general derivations that can be made from dissonance theory; principally, that under conditions specified in this experiment, opinion conformity increases as a function of involvement and discrepancy. This theory, with its consideration of the tension-producing effects of these variables, provides a sound framework for the understanding of some of the dynamics of the process of social influence.

Involvement

Although issue-involvement and response-involvement are not necessarily exclusive, they can be independently varied and should be independently conceptualized. Essentially, an experimenter is said to be using issue-involvement if the mere presentation of the issue is sufficient to elicit concern and interest. Since arousal of issue-involvement is dependent upon a cluster of antecedent factors, which generally cannot be analyzed separately, it tends to be a rather unreliable independent variable.

It seems quite plausible that under such conditions of involvement, *resistance* to change may increase as involvement increases. Hovland and Sherif[3] apparently refer to this kind of issue-involvement when they predict less positive change (and more negative change) as involvement increases. However, the results of a recent unpublished study by Brehm and Lipsher are contrary to the predictions of Hovland and Sherif and accord with the results of the present study. Brehm and Lipsher[4] manipulated involvement using issues which varied in the degree to which *S*s were concerned with them. They found that "The amount of *opinion change* in the direction of the communicator was greater for the issue of high concern than for the issue of low concern" (p. 11).

---

[3] Unpublished manuscript by C. I. Hovland and M. Sherif entitled "Studies in judgment."

[4] Unpublished manuscript by J. W. Brehm and D. Lipsher entitled "Communicator-communicatee discrepancy and perceived communicator trust-worthiness."

The other major class of involvement, response-involvement, denotes the instrumental relationship of a given opinion or response to the achievement of a desired goal. The opinion becomes a means of securing reward, approval, or recognition, of avoiding punishment, or of raising the individual's self-esteem. The involvement is not dependent upon the content of the topic per se, but is imposed upon the topic. The involvement is *created* by conditions specifying the consequences attendant upon task performance. Thus, by imposing involvement on a neutral issue, strength of involvement should be a function of strength of the experimental induction, and should be uncontaminated by differential levels of such antecedent variables as information, support, and commitment. This was a primary consideration in the present interpretation of Festinger's general statement of involvement as dealing with response involvement.

Discrepancy

The present study offers further support for the well documented finding of a positive relationship between extent of discrepancy and total amount of opinion change (Fisher et al., 1956; Fisher & Lubin, 1958; Goldberg, 1954; Harvey et al., 1957; Hovland & Pritzker, 1957). This conclusion remains valid even after controlling for regression artifacts and ceiling effects. Furthermore, its range of generalization has been broadened by the variety of ways in which it has been operationalized from discrepancies in judgments of intelligence, parachutes and dots from photos, to discrepancies in opinions on factual as well as attitudinal issues.

While replication under such diverse conditions increases the reliability of the conclusion, nevertheless the metric of change is difficult to compare across the studies because of their diversity. The measure used in the present study has a number of advantages which recommend its wider use. It is simple to administer and score, yields a description of the affective *structure* of a given opinion, and takes account of subtle changes within the structure. Thus, many of the *S*s in the present study who did not change their "most acceptable" positions showed at least one positive change in one of the components of the opinion. Therefore, measurement of changes in only the "most acceptable" position underestimates the efficacy of the persuasive communication. Similarly, one might predict vulnerability to further influence and propaganda if an individual's latitude of acceptance was extended and/or his latitude of rejection was reduced by an initial communication—even though no change occurred in his "most acceptable" position.

The sensitivity of the composite change index can be increased by use of a scaling procedure (probably Coombsian) which assigns weights on a more empirical, less arbitrary basis. The weighing system should take account of the psychological scale *distance* of each kind of change and of the relative importance of the *quality* of each component of change.

Determining the placement of communications by *S*s' latitudes of acceptance and rejection rather than by their placement at some fixed scale distance from *S*s' "most acceptable" position stems from a suggestion by Hovland and Sherif (see fn. 3). These authors feel that knowledge of the affective structure of a given opinion will aid in the optimal placement of a persuasive communication. The most obvious and rational prediction is that a person is more likely to change and accept a communication which he has previously stated he finds slightly discrepant but "acceptable," than a communication which he previously has found very discrepant and "objectionable." Indeed, the very definition of latitude of rejection implies that there are a number of positions which the individual would reject, and perhaps even respond to by changing negatively. The prediction from dissonance theory of more positive change as a consequence of communications within the latitude of rejection than of those within the latitude of acceptance is not at all obvious.

It has recently been pointed out that as discrepancy becomes extreme, a point is reached where incredulity becomes a factor mediating against opinion change. While Festinger (1957) does not explicitly specify the relationship between incredulity and discrepancy, he would probably predict that dissonance is a monotonically increasing function of discrepancy, regardless of the extremity of the discrepancy. However, with extreme discrepancy, the prepotent dissonance-reducing response should shift from opinion change to one of dissociation, disbelief, or the discrediting of the communicator—even though experimental attempts (as in this study) are made to minimize these reactions. Therefore, dissonance theory would appear to make a prediction similar to that of Osgood and Tannenbaum (1956).

The Communicator

In the present study the communicator was highly positive for all experimental conditions; both attractive to *S* (a friend) and relevant to the issue of discrepancy (an expert on a related issue). This complex definition of positive communicator was intentionally employed in order to firmly "anchor" the position of the communicator.

While it seems obvious that "negative" communicators are generally ineffectual in producing conformity responses, under certain conditions such communicators may induce more positive change than positive communicators. Consider the situation in which an individual is "forced" to comply to a persuasive communication, i.e., to adopt a public stand or to behave in a manner contrary to his own convictions. Dissonance will be created between the sets of cognitions corresponding to his discrepant public and private positions (Festinger, 1957, Ch. 4). Subsequent (real) attitude change, as a mode of reducing the dissonance, will be an inverse function of the degree of justification for the conforming behavior. The more negative the communicator, the less the justification and the greater the dissonance produced by compliance. Thus, a negative communicator could be more influential under conditions of forced compliance than a positive communicator.

Clearly, a whole program of research is necessary in which discrepancy is varied over a wide range under various communicator conditions with several levels of both response and issue-involvement in order to answer the host of theoretical and practical questions concerning the determinants of opinion conformity.

SUMMARY

This research studied the relationship between conformity and (*a*) the extent of the discrepancy between the opinions of a communicator and a recipient and (*b*) the degree of involvement of the recipient. These variables are central to a dissonance theory analysis of the social influence process, as well as to many previous investigations of attitude change. It can be derived from the theory that opinion change increases with increases in both involvement and discrepancy between communicator and recipient.

In order to test these hypotheses, a "before-after" design was used in which 40 pairs of college girls privately gave their opinions about a case of juvenile delinquency. They were exposed to the alleged opinion of their friend (positive communicator), and then gave their opinions a second time. Response-involvement and discrepancy were experimentally manipulated in a 2 × 2 design. Attempts were made to maximize the probability of opinion change as the response to this kind of social influence.

Several different measures of opinion change (including a newly developed one) are consistent in showing that:

1. Highly involved *S*s changed significantly more than *S*s not involved.
2. Opinion change increased significantly as the extent of the discrepancy between communicator and recipient increased.
3. The interaction between these variables was not significant.
4. The maximum dissonance group changed significantly more than the other groups, while the minimum dissonance group consistently changed least. The additional finding of reliable differences between conformers and nonconformers in the relative sizes of their initial latitudes of acceptance and rejection suggests that this may be a useful index of susceptibility to change.

REFERENCES

FESTINGER, L. *A theory of cognitive dissonance.* New York: Row, Peterson, 1957.

FISHER, S., RUBINSTEIN, I., & FREEMAN, R. W. Intertrial effects of immediate self-committal in a continuous social influence situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 200–207.

FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, **56**, 230–238.

GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.

HARVEY, O. J., KELLEY, H. H., & SHAPIRO, M. M. Reactions to unfavorable evaluations of the self made by other persons. *J. Pers.*, 1957, **25**, 393–411.

HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

OSGOOD, C. E., & TANNEBAUM, P. The principle of congruity and the prediction of attitude change. *Psychol. Rev.*, 1956, **62**, 42–55.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

(Received December 12, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE EFFECTS OF INFORMATION ON THE ASSESSMENT OF ATTITUDES BY AUTHORITARIANS AND NONAUTHORITARIANS[1]

MILTON E. LIPETZ[2]
Ohio State University

This study is concerned with the effects of differentially relevant information on the social judgments of authoritarians and nonauthoritarians. Previous research (Crockett & Meidinger, 1956; Scodel & Freedman, 1956; Scodel & Mussen, 1953) has indicated that authoritarians assume that others would respond like themselves to the California F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), whereas nonauthoritarians, while more variable, tend to attribute to their partners F scale scores in the middle of the distribution from which they themselves were selected and thus tend to judge their partners as scoring higher than themselves. In these studies, judgments were made after brief interactions during which topics such as radio, television, and the movies were discussed. Crockett and Meidinger (1956) had their subjects also discuss child-rearing practices, but the amount of information made available through these interactions and its degree of relevance to the judgments remained uncontrolled. Scodel and Freedman (1956) and Christie and Cook (1958) have suggested that the inability of *S*s low in F to be more accurate may be due to the absence of detailed relevant information or to the briefness of interaction.

If *S*s high in F are governed by a stereotype of "sameness," whereas *S*s low in F are not characterized by stereotyped thinking, then the availability of information relevant to a required social judgment (in this case, judgments on the F scale) should result in greater accuracy for the latter group. The literature contains some evidence that nonauthoritarians benefit more from relevant information than do authoritarians (Juul, 1953; Kogan, 1956).

A further question was posed: If *S*s high in F but not those low in F judge others to hold attitudes similar to their own despite contrary relevant information, would measures for which the information was less relevant also reflect this differential behavior? To obtain judgments on a variable for which the information provided was less relevant than for the F scale, a set of value statements was used. Knowledge of a person's attitudes, especially relevant to judgments of his F scale score, should have intermediate relevance to judgments about his values.

It was hypothesized that, given information about relevant sample attitudes of other persons, *S*s low in F should be more accurate than *S*s high in F in differentiating a high F personality from a low F personality. Furthermore, given the same information, *S*s low in F should see their own values as more disparate from those of the high F personality than from those of the low F personality. *S*s high in F should make no such discrimination.

METHOD

Subjects

To select groups high and low in F, 28 items of the Form 40-45 California F scale were administered to 358 male college students. The scale was scored by adding a constant of four to each of the responses. Ninety-six *S*s, 48 in the high group and 48 in the low group, were selected to participate in the experiment. The high F group had total scores ranging from 121 to 148 with an item mean of 4.6. The scores of the low F group ranged from 31 to 85 with an item mean of 2.6. No restrictions were placed on the sample with respect to such factors as age, race, or religious affiliation. An estimate of intelligence was obtained for each *S* by means of the Ohio State Psychological Examination.

Measures

To obtain judgments about values, Wertheimer's six values (1957) were used. These values, each expressed in a single sentence without the value label, are Approval (A), Authoritarianism (F), Hedonism (H), Integrity (I), Other-direction (O), and Status (S). For example, the statement representing Integrity reads: One should satisfy one's own conscience, and make decisions one believes are right, regardless of whether one hurts oneself or others. Status is expressed by: One should act in such a way as to make the most of oneself, in order to earn the respect of others. Each *S* ranked the six values for himself and for the person being judged. Discrepancy scores between self values and judged values were obtained by summing the differences between the two rankings of the six values, yielding a possible range of 0–18 for any one *S*. (For example, if self values were ranked I, S, O, A, H, F, and judged values were ranked F, I, H, A, O, S, the sum of the differences is $1 + 4 + 2 + 0 + 2 + 5 = 14$.)

---

[1] This paper is based on a doctoral dissertation submitted to Ohio State University. The writer wishes to express his sincere appreciation to his advisor, Alvin Scodel, for his encouragement and stimulating criticisms during the entire course of the dissertation.

[2] Now at the University of Colorado.

*S*s were also asked to fill out the F scale for the person being judged.

Source of Information

Mock interview recordings provided *S*s with information upon which to base their judgments. These recordings were made from prepared scripts whose content was derived from the theoretical clusters of Adorno et al. (1950, p. 255). One script was written to portray the high F personality and is hereafter referred to as the high role. The other represented the low F personality and is referred to as the low role. For greater representativeness four different recordings of each role were made with variations in respondents and in spontaneous and nonrelevant comments. The four persons who served as respondents in the high role recordings also served as respondents in the recordings for the low role. In each of the eight recordings the author served as the interviewer. The announced purpose of the interview was to obtain the attitudes of college students on a number of different issues as part of a research program.

In each high role recording the interviewer began by asking the respondent to express his views on juvenile delinquency. In the course of conversation the respondent expressed such opinions as: children are being pampered, children do not respect authority, and people turn to psychiatrists with their troubles instead of bearing them. In response to a question as to whether or not people should try to change the laws of the country, the respondent berated radicals and "low class" people. This interchange is designated hereafter as Information Session I. Following a brief pause the interview continued with the respondent making statements about the mysterious things going on that cannot be understood, his dislike of people who are different, his concern about sexual immorality, his belief that people cannot be changed, the need to control inferior people, and his disgust with the intermingling of races that occurred in Europe during the war. The conversation following the pause is identified as Information Session II. All the information that comprised the high role was consistent with the theoretical clusters upon which the F scale is based. An attempt was made to cover all the clusters, but in the absence of empirical data there was necessarily some subjectivity involved in determining the comprehensiveness of each role.

In each low role recording, none of the information divulged was consistent with these opinion clusters. In response to the question on juvenile delinquency the respondent replied that the problem was complicated, although the instability of society and economic factors contributed to it. At any rate severe punishment would not solve the problem. With regard to changing the laws of the country he felt that, as social conditions changed, laws had to change, although the Bill of Rights should not be diluted. During Information Session II the respondent expressed concern over social conditions, segregation, and the lack of understanding among peoples of different nations. With respect to interracial marriage he felt that this was the concern of the persons involved.

In both the low and high roles, the opinions expressed during Session II were more forceful than those offered in Session I. The adequacy of each role was informally passed on by six judges who were familiar with the theoretical clusters of the F scale. Since both roles were written to correspond to extreme positions, it is assumed that the higher a predicted score is for the high role and the lower a predicted score is for the low role, the greater the accuracy of judgment.

Procedure

From one to seven *S*s were present in any one session. They were informed that they had been selected to participate in a study investigating the value of recordings for determining the nature of social interactions. Any one *S* was exposed to only one of the eight recordings, each of which was heard by an equal number of *S*s, half high in F and half low in F. Prior to the playing of the recording, *S* was asked to rank the six Wertheimer values in order of their importance to him. The recording was then played through the end of Information Session I. Immediately thereafter, *S* was requested to fill out the F scale and to rank the six values as he thought the respondent on the recording would have answered. The recording was then continued through the end of Information Session II, after which repeated estimates of the interviewee's responses were obtained.

RESULTS

An analysis of variance of F scale judgments among records within each role by subject group and information session failed to reveal any differences significant at the .05 level. Data from the four recordings within each role were therefore combined in further computations.[3]

High Role versus Low Role

The mean estimate scores for the high and low roles are presented in Table 1. In order to verify the assumption that respondents in the high role recordings would be perceived as

---

[3] Since intelligence could be related to the ability to use information for prediction purposes, partial correlations among intelligence, self F score, and predicted F score were computed. For this sample, intelligence is not a crucial factor for the prediction of F scores of other persons.

TABLE 1

COMPARISON OF MEAN ESTIMATED F SCORES FOR HIGH AND LOW ROLES

Role on Record	Subject Group	Information Session	N	Mean	SD	t	df^a	p
High	Total	I	48	155.1	12.60	12.38	77	<.001
Low	Total	I	48	110.8	21.43			
High	Total	II	48	168.4	12.37	18.17	74	<.001
Low	Total	II	48	103.0	12.81			
High	High	I	24	157.7	11.94	9.39		<.001
Low	High	I	24	123.6	12.67			
High	High	II	24	167.8	11.64	13.77		<.001
Low	High	II	24	117.8	12.94			
High	Low	I	24	152.4	12.69	10.73	39	<.001
Low	Low	I	24	98.0	20.79			
High	Low	II	24	168.9	13.04	16.97	43	<.001
Low	Low	II	24	88.3	18.67			

[a] Degrees of freedom equal $N_1 + N_2 - 2$. Where variances are unequal,

$$df = \frac{n\,\frac{(s_1^2}{N_1} + \frac{s_2^2)^2}{N_2}}{\frac{(s_1^2)^2}{N_1}\,\frac{1}{N_1+1} + \frac{(s_2^2)^2}{N_2}\,\frac{1}{N_2+1}} -2$$

(from Walker & Lev, 1953, p. 158)

scoring higher in F than those in the low role recordings, the roles were compared for the total subject group for each information session. The high role was, on the whole, differentiated from the low role for both information sessions. A comparison of *S*s high in F judging the high role and *S*s high in F judging the low role indicates that *S*s high in F were able to differentiate between high and low roles for both information sessions. A similar comparison for *S*s low in F resulted in the same finding.

The mean difference between the high role judgment and the low role judgment for *S*s high in F was compared with that for *S*s low in F. Both after Information Session I and after Session II, *S*s low in F perceived greater differentiation between the two roles than did *S*s high in F (I: $t = 3.25$, $p < .01$; II: $t = 5.12$, $p < .001$). The first hypothesis is thus supported.

High Subjects versus Low Subjects

*S*s high in F did not differ significantly from *S*s low in F in their judgments of the high role recordings following Information Session I ($t = 1.46$) or Information Session II ($t = .30$). On low role recordings, however, *S*s low in F perceived the role as significantly lower than did *S*s high in F following Information Session I ($t = 5.05$, $p < .001$) and Information Session II ($t = 6.24$, $p < .001$). The difference between *S*s high and low in F on predictions for the low role was significantly greater than that for the high role after Information Session I ($t = 3.25$, $p < .01$) and after Information Session II ($t = 4.75$, $p < .001$). Following both information sessions *S*s low in F thus perceived the low role more accurately than did *S*s high in F. The two groups of *S*s did not differ in their perceptions of the high role.

Information Session I versus Information Session II

A comparison of Information Session I and Information Session II for *S*s high in F who judged the high role reveals that predictions were significantly more accurate following Information Session II ($t = 2.90$, $p < .01$). For the low role, *S*s high in F were not significantly more accurate after Information Session II ($t = 1.53$). Similar results were obtained for *S*s low in F; the high role was perceived more accurately after Information Session II than after Information Session I ($t = 4.35$, $p < .001$), but there was no significant difference with respect to the low role ($t = 1.67$).

Perhaps these findings might be influenced by ceiling effects. The highest judgment one could make for the high role was a score of 196. The lowest judgment one could offer for the low role was a score of 28. The more accurate a prediction was after Information Session I, the less it could be improved after Information Session II because of these set limits. To remove this inequity, improvement was rated relative to the possible improvement that could have occurred between the end of Information Session I and the end of Information Session II. Improvement was greater for *S*s low in F than for *S*s high in F for the high role ($t = 2.14$, $p < .05$). Improvement on the low role did not differ for *S*s high and low in *F* ($t = 1.07$).

TABLE 2

MEAN SUMS OF DIFFERENCES BETWEEN SELF-RANKINGS OF VALUES AND RANKINGS OF ASCRIBED VALUES

Role on Record	Subject Group ($N = 24$)	Information Session	Mean	*SD*	*t*	*df*	*p*
High	High	I	8.0	4.04	4.85	36	<.001
High	Low	I	12.8	2.44			
Low	High	I	10.1	4.98	.51		—
Low	Low	I	9.4	3.76			
High	High	I	8.0	4.04	1.55		—
Low	High	I	10.1	4.98			
High	Low	I	12.8	2.44	3.58	40	<.001
Low	Low	I	9.4	3.76			
High	High	II	9.6	4.16	4.00	38	<.001
High	Low	II	13.6	2.38			
Low	High	II	8.0	3.74	.55		—
Low	Low	II	8.7	4.42			
High	High	II	9.6	4.16	1.35		—
Low	High	II	8.0	3.74			
High	Low	II	13.6	2.38	4.68	37	<.001
Low	Low	II	8.7	4.42			

Discrepancy between Own and Ascribed Values[4]

The means of the sum of differences between self-rankings of values and rankings of values ascribed to respondents in the recordings are presented in Table 2. *S*s low in F saw a greater discrepancy between their own values and those of respondents in the high role than did *S*s high in F following each information session. The two groups did not differ significantly in their discrepancy scores for the low role after either information session. *S*s high in F did not discriminate between ascribed values for the high and low roles following either information session, whereas *S*s low in F saw a greater discrepancy between self-values and values of the high role than between self-values and values ascribed to the low role after each information session. The second hypothesis is supported.[5]

DISCUSSION

The data on judgments of F scale scores show that *S*s high in F are able to make social discriminations when information is sufficiently relevant. It is not believed that the present finding represents evidence contradictory to the findings of previous studies (Crockett & Meidinger, 1956; Scodel & Mussen, 1953) that *S*s high in F distort information to satisfy a need to see others like themselves. But there seem to be definite limits within which distortion can occur. When information is as direct and obvious as in this study, *S*s high in F take it into account. However, their ability to discriminate is not as great as that of *S*s low in F. The advantage of *S*s low in F is due to their ability to perceive the low role more accurately. Even when information is highly relevant, the need of *S*s high in F to see others like themselves persists to some degree.

It has been noted that the present findings indicate that *S*s low in F differentiated between the two roles even more sharply than *S*s high in F. In previous studies (Crockett & Meidinger, 1956; Scodel & Freedman, 1956) *S*s low in F tended to place others in a middle category. The results obtained here indicate that *S*s low in F, given relevant information, are able to make accurate social judgments. Yet Crutchfield's (1954) suggestion that *S*s low in F may have an "other" stereotype (rather than the "same" stereotype of *S*s high in F) still seems valuable. An inspection of the raw data reveals that it is a rare *S* low in F who judges the scores for the low role as lower than his own (two of 24 after Information Session I; five of 24 after Information Session II). Similarly, Rabinowitz (1956) found a marked tendency among nonauthoritarians to estimate F scale scores for the typical college student as higher than they themselves possessed.

The inability of *S*s low in F to benefit from the additional information of Session II could be due to the role script itself. It could also be due to their own possible holding of irrational stereotypes. If, for the most part, they see others as having higher F scale attitudes than their own, they have already set an irrational lower limit for the low role; hence, they are unable to use the additional information.

---

[4] In a comparison of differences in self ranking of values for the two groups, *S*s high in F ranked authoritarianism significantly higher than did *S*s low in F (chi square value 7.19, $p < .01$). *S*s low in F ranked Integrity as significantly higher than did *S*s high in F (chi square value 7.43, $p < .01$).

[5] To extend the dimension of relevance, judgments were also sought on variables (height and weight) for which the information provided was in no way relevant. Analyses of discrepancies between self height and weight and judged height and weight resulted in generally negative findings. The one significant comparison indicated that *S*s low in F judged the low role as more similar to themselves in height and weight than the high role. For a fuller discussion, the reader is referred to the author's dissertation (Lipetz, 1958).

The analyses involving value discrepancies between self and judgments offer some interesting findings. There can be little doubt that the information on the recordings was relevant for the ranking of values. The relevance, however, was not as direct as for the F scale judgments, since the scripts were written specifically with F content in mind. With reality factors less compelling, *S*s low in F were still sensitive to differences between the roles. The data obtained from *S*s high in F, however, are consistent with the Scodel-Mussen (1953) assertion that the person high in F maintains his security by assuming that the values of others are similar to his own.

In conclusion, it seems that both *S*s high and low in F are governed to some extent by stereotyped thinking, albeit of different types.

SUMMARY

This study was concerned with the effects of differentially relevant information on social judgments of authoritarians and nonauthoritarians. After listening to standard recorded interviews, the content of which was derived from the theoretical clusters of the F scale, *S*s made judgments about the respondent's F scale attitudes and his values.

It was hypothesized that *S*s low in F are better able to use information for making accurate judgments. On the whole, results confirmed the hypothesis, although there was some evidence that *S*s low in F, as well as those high in F, tended toward stereotyped thinking.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

CHRISTIE, R., & COOK, PEGGY. A guide to published literature relating to the authoritarian personality through 1956. *J. Psychol.*, 1958, **45**, 171–199.

CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.*, 1956, **53**, 378–380.

CRUTCHFIELD, R. S. Social psychology and group processes. *Annu. Rev. Psychol.*, 1954, **5**, 171–202.

JUUL, K. D. Authoritarian personality in relation to teachers' attitudes toward child behavior. Unpublished doctoral dissertation, Wayne Univer., 1953.

KOGAN, N. Authoritarianism and repression. *J. abnorm. soc. Psychol.*, 1956, **53**, 34–37.

LIPETZ, M. E. The effects of information on the assessment of attitudes by authoritarians and nonauthoritarians. Unpublished doctoral dissertation, Ohio State Univer., 1958.

RABINOWITZ, W. A note on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **53**, 384–386.

SCODEL, A., & FREEDMAN, MARIA L. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.

SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WERTHEIMER, M. Values in person cognition. In USAF Behav. Sci. Conf. Rep., Univer. New Mexico, 1957, *Decisions, values, and groups*. New York: Pergamon Press, in press.

(Received January 7, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SOCIALLY MEDIATED REDUCTION IN EMOTIONAL RESPONSES OF YOUNG RHESUS MONKEYS[1]

WILLIAM A. MASON[2]
University of Wisconsin

CHILDREN frequently become upset and fearful when alone in strange surroundings (Arsenian, 1943; Jersild & Holmes, 1935), and nonsystematic observations of young monkeys and chimpanzees suggest that isolation in an unfamiliar situation produces similar effects in these nonhuman primates (Finch, 1942; Foley, 1934; Haslerud, 1938; Köhler, 1927; McCulloch, 1939).

Arsenian (1943) has demonstrated experimentally with preschool children that emotional distress induced by placing the child alone in a novel room is substantially reduced if a familiar adult is present. Similarly, the presence of familiar social stimuli has been observed to mitigate the emotional reactions of nonhuman primates to novel or frustrating situations (Haslerud, 1938; McCulloch, 1939; Watson, 1914).

Although it is reasonable to assume that the effectiveness of a specific social stimulus in reducing disruptive emotional responses is dependent upon the nature and extent of previous contacts with it, this assumption has apparently not been subject to systematic investigation. For most immature primates, the mother is no doubt highly effective in reducing disruptive emotional responses, but there is evidence that other adults (Arsenian, 1943) or children (Wright, 1942) have a similar effect. Interpretation of this finding is difficult, however, in view of demonstrated interaction between experience with the mother and subsequent responses to other social stimuli (Bishop, 1951).

The present subjects (*S*s) provided a large measure of control over possible interactions of this sort. They were young monkeys, born in the laboratory and raised under conditions which, with the specific exceptions described below, limited social experience with other monkeys to visual and auditory contacts with age-peers. The laboratory environment provided everything essential to maintain the *S*s in good health, but in a routine and impersonal manner. Thus their social history severely restricted opportunity for the development of complex social interactions and expectancies, particularly in regard to other monkeys.

The present researches investigated the influence of various social stimuli on the emotional responses of these *S*s by testing them in isolation and in the presence of familiar and unfamiliar social stimuli.

EXPERIMENT 1

Method

Subjects. The *S*s were 12 young rhesus monkeys 250 to 263 days of age. They were born in the laboratory, separated from their mothers within 18 hrs. after birth, and subsequently housed in individual living cages which allowed them to see and hear other young monkeys, but restricted other forms of social contact. Opportunity for unrestricted social interaction was limited to 60 daily 3-min. free-situation tests with the same partner completed approximately one week before the present experiment began. These tests were conducted in the cubicle subsequently described. Each *S* had had five additional 3-min. periods alone in the cubicle in the week preceding the present experiment.

Apparatus. The test chamber was a 6 ft. by 6 ft. by 6 ft. plywood cubicle illuminated from above by a 150 w. bulb. Two opposing walls of the chamber contained one-way observation panels 15½ in. by 11½ in. and in each of the other walls was a large hinged door fitted with a smaller sliding door through which the *S* entered and left the chamber. Only one sliding door was used in the present experiment. Near the wall opposite this door was a Plexiglas detention cage measuring 18 in. high, 12 in. wide, and 12 in. deep. The floor of the chamber was marked off into 2 ft. squares.

In addition to the detention cage, the only other objects within the cubicle were a red wooden cylinder 3 in. in diameter and 3½ in. high, a polished brass knob within a small box, enclosed on all sides but one, an unpainted wooden dumbbell 9 in. long, and a "tree" consisting of a wooden cylinder 3¼ in. in diameter and 15½ in. high into which several "limbs" of brass rodding were fixed. These objects were provided as incentives for visual and manual exploration. Two objects

---

[1] Support for this research was provided through funds received from the Graduate School of the University of Wisconsin, Grant G6194 from the National Science Foundation, and Grant M-722 from the National Institutes of Health.

[2] Now at Yerkes Laboratories of Primate Biology, Orange Park, Florida.

were placed near each of two walls on opposite sides of the room.

Procedure. At the beginning of each session the *S* was brought to the test area in a carrying cage and placed outside the cubicle in front of the sliding door. A trial was started by raising this door and introducing the *S* into the room. The 3-min. trial period was divided into 12 intervals each 15 sec. in length and the intervals were indicated by a Veeder-Root counter activated by a recycling timer. The *S* was observed through the one-way vision screens and its behavior recorded on a time-ruled check sheet. The recorded behaviors included the following emotional responses described by Foley (1934, 1935): (*a*) *Crouching-clasping.* Assumption of a crouching posture with limbs close to the body and head below the level of the shoulders and/or clasped its fur, limbs, or other body parts with hands or feet. (*b*) *Stereotyped movements.* Fixed, repetitive movements not resulting in locomotion. Rocking or swaying of the upper body were most frequently observed, but rubbing or stroking of body parts or contiguous regions of the chamber were also included in this category. (*c*) *Sucking.* Insertion into the mouth of fingers, toes, or other body parts. (*d*) *Vocalization.* Two forms of vocalization were recorded, a high-pitched, moderately intense "coo," and an intense scream or screech. A record was also obtained of *approaches to the detention cage*, *approaches to the objects* (the criterion of approach in each instance being 6 in.), and *contacts* with the objects.

Occurrence of any of the foregoing behaviors was recorded at the appropriate time interval on the check sheet, and scores for behavior categories were based on the number of 15-sec. intervals in which these behaviors occurred. No interval could contribute more than one to the total score for any behavior category regardless of the number of occurrences of the behavior within that interval. Locomotor activity was measured by advancing a counter whenever the *S* entered a square up to the midline of its body.

To provide an estimate of the reliability of the method, two *O*s were present and independently recorded responses for four *S*s on a total of 26 trials. Reliability measures were not determined for locomotion scores, and vocalization occurred too infrequently in these *S*s to warrant calculation of reliability coefficients. For all other measures, product-moment correlation coefficients ranged from .94 to .99.

The effects of three stimulus conditions were investigated: Each *S* was observed when the detention cage was empty (empty = Em), when the cage contained the partner that had been paired with the *S* in the 90 free-situation tests preceding the present experiment (partner = Pa), and when the cage contained an unfamiliar monkey of approximately the same age as the *S* (stranger = St). Stimulus conditions were presented in one of three fixed orders and all conditions appeared once before any condition was repeated. Each condition was presented five times.

Results

Total scores for each behavior category and their percentage equivalents are given for test conditions in Table 1. All emotion measures increased under the Em condition, but the effect of test conditions was statistically significant only for crouching-clasping and stereotyped movements. These measures were significant at the .01 level as determined by the Friedman nonparametric analysis of variance (Siegel, 1956).

TABLE 1

RESPONSES TO PARTNER, STRANGER, AND EMPTY STIMULUS CONDITIONS

	Partner		Stranger		Empty	
	Freq.	%	Freq.	%	Freq.	%
Crouching-clasping	265	37	264	37	409	57
Stereotyped movements	140	19	127	18	226	31
Sucking	68	9	50	7	86	12
Vocalization	15	2	10	1	31	4
Composite score	488	17	451	16	752	26
Aproach cage	601	83	611	85	425	59
Approach objects	223	31	265	37	239	33
Object-contact	182	25	205	28	182	25
Mean locomotion/trial	24.6		27.7		22.9	

Emotion scores were additively combined to yield a composite emotion score. The percentage of composite emotion scores were 17, 16, and 26 for the Pa, St, and Em conditions respectively, and the difference among conditions was significant at the .01 level. Comparisons between individual test conditions with the Wilcoxon test (Siegel, 1956) indicated that the Em condition differed from the Pa and St conditions at the .01 level. The difference between the Pa and St conditions was not statistically significant.

Approaches to the cage dropped from approximately 84% under the Pa and St conditions to 59% under the Em condition, and the effect of stimulus conditions on this measure was significant at the .01 level as determined by the Friedman test. Locomotion scores were also lowest under the Em condition, but differences among conditions were not statistically significant. There was no indication that approaches to objects or contacts with objects were systematically related to stimulus conditions.

The present experiment demonstrated that presence of a social stimulus produced a reduction in emotional responsiveness, but provided no evidence that the specific characteristics of the social stimulus and *S*s' previous experience with it were effective variables. Therefore, another experiment was conducted that used a wider range of social stimuli.

EXPERIMENT 2

Method

Subjects. The six pairs of rhesus monkeys used in Experiment 1 served again as *S*s. Their ages ranged from 439 to 536 days. In connection with another experiment, six of the monkeys had been housed with their partners and had been living together for 18 weeks when the present experiment began. For the other *S*s, living conditions were as previously described. Following Experiment 1, opportunities for social contact with the partner were provided to all *S*s in 20 social preference tests, each 2 min. long and conducted in the test

TABLE 2

RESPONSES TO PARTNER, STRANGER, ADULT, RABBIT, AND EMPTY STIMULUS CONDITIONS

	Partner		Stranger		Adult		Rabbit		Empty	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Crouching-clasping	137	19	171	24	229	32	269	37	276	38
Stereotyped movements	70	10	72	10	124	17	143	20	124	17
Sucking	65	9	54	7	91	13	106	15	103	14
Vocalization	2	0	12	2	24	3	36	5	14	2
Composite score	274	9	309	11	468	16	554	19	517	18
Approach cage	671	93	654	91	497	69	557	77	581	81
Approach objects	113	16	101	14	173	24	108	15	97	13
Object-contact	100	14	92	13	141	20	101	14	84	12
Mean Locomotion/trial	36.0		32.6		26.3		24.2		29.1	

cubicle previously described, and in tests of food-sharing and communication. Only the tests of social preference permitted unrestricted social interaction.

Apparatus. The cubed room was again used and the experimental situation was unchanged, except that vertically striped panels were suspended from two walls of the test chamber and the original objects were replaced by four new stimulus objects: a coffee can, a grey rubber hemicylinder, a small aluminum-painted block with dowels protruding from it, and an orange and red "tree." The dimensions of these objects were approximately the same as those described in Experiment 1.

Procedure. Testing and recording procedures and behavioral measures were the same as those used in Experiment 1, except for the addition of two stimulus conditions bringing the total number to five: Each *S* was observed when the detention cage was empty (Em), when the cage contained the partner (Pa), an unfamiliar monkey of the same age (St), an unfamiliar adult (Ad), and a large albino rabbit (R). The *S*s had never before seen a rabbit and had seldom if ever seen adult monkeys. All stimulus conditions appeared once before any condition was repeated and each condition was presented five times.

Results

The data were first examined for differential effect of living arrangements on responses to the Pa and St conditions. The *S*s living continuously with the partner gave no evidence that the partner was more effective in reducing emotional responses than a stranger of the same age. For all *S*s differences between the Pa and St conditions were slight, and there were no significant differences between paired and solitary *S*s in response to these stimulus conditions. In view of these findings the data from all *S*s were pooled for further analysis.

Total scores for each behavior category and their percentage equivalents are given for test conditions in Table 2. Although emotion scores were lower under the Pa and St condition, the effect of stimulus conditions was significant only for crouching-clasping. Emotion scores were additively combined to yield a composite emotion score. The percentage of composite emotion scores for the Pa, St, Ad, R, and Em conditions were 9, 11, 16, 19, and 18, respectively, and the effect of stimulus conditions on this measure was significant at the .01 level as determined by the Friedman test. Comparisons between individual stimulus conditions with the Wilcoxon test indicated that the Pa condition differed from the Ad, R, and Em conditions at the .02 level. The difference between the St and R conditions was also significant at the .02 level. All other differences were nonsignificant.

The percentage of approaches to the cage were 93, 91, 81, 77, and 69 for the Pa, St, Em, R, and Ad conditions, respectively, and the difference among conditions was significant at the .01 level as determined by the Friedman test. Comparisons between stimulus conditions indicated that the Ad condition differed from the Pa, St, and Em conditions at the .05 level as determined by the Wilcoxon test. All other differences were nonsignificant. Mean locomotion scores were highest in the Pa and St conditions, but differences among conditions were not significant for this measure or for approaches to objects and contacts with objects.

Approximately 6 months separated Experiments 1 and 2. The absolute magnitude of composite emotion scores in Experiment 2 was somewhat lower than for comparable stimulus conditions in Experiment 1, suggesting that the intervening experience in the test chamber and/or maturational changes had produced some reduction in the general level of emotional responsiveness. Despite this change, however, a rank order correlation performed on total individual composite emotion scores yielded a coefficient of .74 ($p =$ .02), indicating that relative to other members of the group, individual response characteristics were stable and persistent.

DISCUSSION

Previous observations that social stimuli may function as a source of security and a means of mitigating emotional distress in young primates are fully supported by the present results. The assumption that the familiarity of the social stimulus is an important determinant of the capacity of the stimulus to reduce emotional distress is confirmed by the finding that the level of emotional responses was lower in the presence of monkeys the same age as the *S* than in the presence of an unfamiliar adult monkey or rabbit. This conclusion must be qualified, however, in view of the failure to demonstrate unequivocally that the monkey with which the *S* had had the greatest amount of previous contact was more effective than an unfamiliar monkey of the same age. We possess unpublished data showing that the capacity for social discrimination is highly developed in monkeys of this age, and it is unlikely that failure to discriminate can account for the apparent equivalence of the familiar partner and the strange age-peer. A more acceptable hypothesis is that as the result of visual presence of young monkeys in the *S*'s home environment, all young monkeys had acquired equivalent status as familiar social stimuli in this situation. The effectiveness of the strange age-peer may also have been enhanced, however, as a result of the *S*'s experience with the familiar partner.

Although the present experiments did not investigate the effectiveness of familiar nonsocial stimuli, the research of Harlow and Zimmermann (1958) indicates that equivalent exposure in the living cage to two stimuli, one a preferred social substitute and the other nonpreferred, does not produce equivalent effects. Their research clearly demonstrates that the capacity of a stimulus to mitigate emotional distress is influenced by the specific characteristics of the stimulus in addition to the amount of previous exposure to it.

These considerations, and the obvious similarity of young and adult monkeys, suggest that the present *S*s should have reacted differently to the adult monkey than to the rabbit or the empty cage. Further information as to the possible nature of these differences is provided by Hebb's (1946) theoretical analysis of fear behavior. In his discussion of fear of strangers in young chimpanzees, Hebb proposes that a stimulus which displays sufficient similarity to familiar stimuli to arouse habitual processes of perception, but is also different in certain respects, will produce emotional disturbance. Thus, the adult monkey which was highly similar to the young monkeys with which these *S*s were familiar, but also differed conspicuously from young monkeys in many respects, should have produced emotional disturbance. Although composite emotion scores were actually somewhat lower in the presence of the adult monkey than in the presence of the rabbit or empty cage, the hypothesis that the adult monkey was a source of emotional disturbance is supported by the finding that approaches to the cage were lowest when the adult monkey was present and by the observation that stimulus-directed social responses including threat reactions, lipsmacking, and grimaces (usually interpreted as indicative of fear or submission) were most frequent for this stimulus condition. These responses were recorded systematically but were not included in the presentation of results because they are characteristically elicited by and directed toward social stimuli (Bernstein, 1959; Carpenter, 1942; Chance, 1956), in contrast to the composite emotion measures which occur in the absence of a discrete eliciting stimulus (Bernstein, 1959). Threat reactions, lipsmacking, and grimaces occurred approximately twice as frequently in the presence of the adult monkey as compared with any of the other stimulus conditions, and significant differences were obtained between this stimulus condition and the rabbit and empty cage. These results suggest that the adult monkey had an effect on the emotional responses of these young monkeys similar to that evoked in children and young chimpanzees by strange persons (Hebb & Riesen, 1943; Jersild & Holmes, 1935).

Contemporary theoretical discussions of social development of the human child emphasize the fundamental importance of nurturant activities, particularly feeding experiences, in the development of social attitudes and patterns of social interaction (Dollard & Miller, 1950; Mussen & Conger, 1956; Sears, Whiting, Nowlis, & Sears, 1953). It is highly improbable that the effectiveness of the partner and the strange age-peer in reducing behavior indicative of emotional distress was in any manner derived from or dependent upon nurturant activities. The *S*'s only experience with other monkeys prior to Experiment 1 was confined to 90 brief trials with a single partner and the restricted contacts provided in the living area. From birth, all major caretaking acts were performed routinely and impersonally by laboratory personnel. Although performance of nurturant functions might substantially enhance the capacity of a social stimulus to reduce emotional disturbance, it is clear from the present results that it is not a necessary antecedent of this effect.

SUMMARY

Two studies investigated the effect of various social stimuli upon emotional responsiveness of 12 young rhesus monkeys raised in the laboratory in the visual presence of other young monkeys. Experiment 1 compared the effect of three stimulus conditions: A familiar partner, an unfamiliar age-peer, and the empty detention cage, and Experiment 2 extended the range of test stimuli to include an adult monkey and an albino rabbit.

In both experiments responses indicative of emotional disturbance were less frequent in the presence of the partner and the unfamiliar age-peer. These results suggest that the nature of *S*s' previous social experience was an important determinant of stimulus effectiveness in reducing emotional distress. The capacity of a social stimulus to mitigate emotional disturbance did not appear to be dependent upon nor derived from feeding or other nurturant experiences. Evidence was presented suggesting that the adult monkey was a source of emotional disturbance, producing an effect similar to that observed in young children and chimpanzees in the presence of strangers.

REFERENCES

ARSENIAN, JEAN M. Young children in an insecure situation. *J. abnorm. soc. Psychol.*, 1943, **38**, 225–249.

BERNSTEIN, S. The development of emotional responses to complex stimuli in young rhesus monkeys. Unpublished master's thesis, Univer. of Wisconsin, 1959.

BISHOP, B. M. Mother-child interaction and the social behavior of children. *Psychol. Monogr.*, 1951, **65**, (11, Whole No. 328).

CARPENTER, C. R. Sexual behavior of free ranging rhesus monkeys (*Macaca mulatta*): I. Specimens, procedures, and behavioral characteristics of estrus. *J. comp. Psychol.*, 1942, **33**, 113–142.

CHANCE, M. R. A. Social structure of a colony of *Macaca mulatta*. *Brit. J. Anim. Behav.*, 1956, **4**, 1–13.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

FINCH, G. Chimpanzee frustration responses. *Psychosom. Med.*, 1942, **4**, 233–251.

FOLEY, J. P., JR. First year development of a rhesus monkey (*Macaca mulatta*) reared in isolation. *J. genet. Psychol.*, 1934, **45**, 39–105.

FOLEY, J. P., JR. Second year development of a rhesus monkey (*Macaca mulatta*) reared in isolation during the first eighteen months. *J. genet. Psychol.*, 1935, **47**, 73–97.

HARLOW, H. F., & ZIMMERMANN, R. R. The development of affectional responses in infant monkeys. *Proc. Amer. Phil. Soc.*, 1958, **102**, 501–509.

HASLERUD, G. M. The effect of movement of stimulus objects upon avoidance reactions in chimpanzees. *J. comp. Psychol.*, 1938, **25**, 507–528.

HEBB, D. O. On the nature of fear. *Psychol. Rev.*, 1946, **53**, 259–276.

HEBB, D. O., & RIESEN, A. H. The genesis of irrational fears. *Bull. Canad. psychol. Ass.*, 1943, **3**, 49–50.

JERSILD, A. T., & HOLMES, FRANCES B. *Children's fears*. New York: Teachers Coll. Columbia Univer., 1935.

KÖHLER, W. *The mentality of apes*. New York: Humanities Press, 1927.

MCCULLOCH, T. L. The role of clasping activity in adaptive behavior of the infant chimpanzee: III. The mechanism of reinforcement. *J. Psychol.*, 1939, **7**, 305–316.

MUSSEN, P. H., & CONGER, J. J. *Child development and personality*. New York: Harper, 1956.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. psychol. Monogr.*, 1953, **47**, 135–236.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

WRIGHT, M. E. Constructiveness of play as affected by group organization and frustration. *Character & Pers.*, 1942, **11**, 40–49.

WATSON, J. B. *Behavior: An introduction to comparative psychology*. New York: Holt, 1914.

(Received January 12, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

DIRECT AND VICARIOUS REDUCTION OF HOSTILITY[1]

MILTON E. ROSENBAUM
State University of Iowa

AND RICHARD DECHARMS
Washington University

OBSERVATION of the responses of another to various environmental stimuli may at times affect the response system of the observer as if he had carried out the behavior himself. Phenomena of this type are described as falling under the rubric of vicarious experience (deCharms & Rosenbaum, in press). A minimal amount of research attention has been given to vicarious phenomena, although many students of the social scene claim increasing participation in this mode of experience in contemporary society.

The present report examines the effectiveness of vicarious experience in reducing hostility toward an instigator and at the same time tests hypotheses related to the role of communication in the reduction of aggressive responses.

The theory of frustration and aggression as elaborated by Dollard, Doob, Miller, Mowrer, and Sears (1939) posits that the blocking of goal directed activity leads to the occurrence of aggression of some kind and in some degree. They further postulate in a dependent definition of aggression that aggression "reduces the secondary, frustration-produced instigation, and leaves the strength of the original (nonaggressive) instigation unaffected" (Dollard et al., 1939, p. 11). The statement concerning the reduction of the secondary instigation represents the catharsis hypothesis. More specifically, the catharsis hypothesis reads: "The occurrence of any act of aggression is assumed to reduce the instigation to aggression" (p. 50).

Dollard et al. present minimal evidence in support of the hypothesis. Several later investigators have treated the catharsis hypothesis in contexts involving displacement of aggression. Reduction of experimentally aroused hostility has been studied most frequently as a function of aggression expressed toward objects other than the original instigating agent, i.e., a TAT card (Feshbach, 1955) or associates (French, 1944; Thibaut, 1950; Pepitone & Reichling, 1955). In general, support for the hypothesis was obtained in these studies, although as Berkowitz (1958) indicates, alternative explanations are available in each case.

In two experiments involving expression of aggression toward the instigator himself, Thibaut and Coules (1952) obtained equivocal results related to the catharsis hypothesis. Their basic design involved a situation where, in the course of interaction, *S* was personally insulted in a note sent by a paid participant. Half of the *S*s were allowed to respond immediately, while for the other half the interaction was interrupted at that point. A measure of residual hostility gave results in the direction predicted by the catharsis hypothesis, i.e., those *S*s who were permitted to communicate their hostility directly to the aggressor showed less residual hostility.

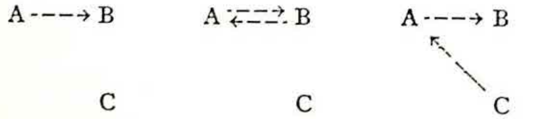
As Thibaut and Coules (1952) pointed out, this finding may have been due to increased frustration produced by the blocking of a hostile response in the interrupted group rather than to a genuine catharsis effect. Experimental evidence was obtained supporting this possibility. In addition, there was no indication within the communication group that the amount of expressed aggression was related to the amount of reduction in hostility, as would have been predicted by the catharsis hypothesis. An implicit assumption of equal instigation is necessary in the examination of this relationship. Evaluation of their own data by Thibaut and Coules indicates that this assumption was not fulfilled.

---

[1] This paper is based on work carried on while the authors were in attendance at the Behavioral Sciences Research Conference at the University of New Mexico during the summers of 1957 and 1958. This conference was supported by the Behavioral Sciences Division, Air Force Office of Scientific Research, under Contract AF 49(6380-33). A portion of the final data analysis was carried out under Contract Nonr-816(11) between the Office of Naval Research and the Small Groups Laboratory, Social Science Institute at Washington University. The authors are indebted to I. E. Farber for a critical reading of the manuscript.

By introducing a third person into an experimental situation similar to that of Thibaut and Coules, it is possible to gain further insight into the dynamics underlying their results and also to explore conditions under which vicarious experience is successful in modifying responses.

Diagrammatically, three experimental conditions can be represented as follows:



1. No Communication (No-Com) 2. Communication (Com) 3. Vicarious Communication (Vi-Com)

In these diagrams the initial aggressor is always represented by A, the *S* who is aggressed against is represented by B, and the potential "vicar" is represented by C. The arrows represent aggressive attacks, or, in the communication situation, the possibility for B to retaliate against A by means of an aggressive communication to him. Situations 1 and 2 are, in effect, those investigated by Thibaut and Coules with the addition of the bystander, C. Situation 3 represents a vicarious setting in which C acts as a "vicar" for B.

An important variable to be accounted for in this model is the strength of the original instigation to aggression. The actual insulting attack, or in terms of the Dollard et al. (1939) formulation, the degree of interference, can be constant across experimental conditions, but individual reactions to the attack may vary. Since instigation to aggression is conceived of as being an interference with goal directed activity, strength of instigation to aggression will be a function of the strength of the instigation to the activity that is interfered with. In the experiment to be described, the activity is the presentation of information about oneself. This task may be of varying importance to the person in the situation, depending on his level of self-esteem. The person with low self-esteem who feels socially inadequate can be expected to be more concerned with presenting an acceptable impression and receiving favorable evaluation than the person with high self-esteem; and, therefore, the attack on the self-picture of the person low in self-esteem may engender more frustration and hence more aggression.

The experiment to be reported was designed to study the residual hostility shown by *S*s after experiencing one of the three conditions in the model. Specifically, the hypotheses to be tested are:

Hypothesis I. A person will manifest less residual hostility after an aggressive attack on him if he is given an opportunity to communicate directly to the aggressor than if such an opportunity is denied.

Hypothesis II. A person will have less residual hostility after an aggressive attack on him if someone else within his awareness attacks the aggressor than if no opportunity for him or anyone else to attack is presented.

The effectiveness of the vicarious situation for reducing hostility will depend upon characteristics of the vicarious attack and upon characteristics of the vicar. Despite these factors one would predict that direct opportunity for attack should be more effective than a vicarious attack in reducing hostility. In effect, we are presenting as a tentative hypothesis the contention that the amount of residual hostility in the vicarious situation should be intermediate between the amounts in the other two situations.

Hypothesis III. A person with low self-esteem will manifest the effects predicted in Hypothesis II and III to a greater extent than will a person with high self-esteem.

METHOD

Subjects

Fifty-six male summer session students at the University of New Mexico participated in the experiment. The *S*s were paid for their two hours of participation. Four of these *S*s spontaneously indicated suspicion of deception and were therefore not considered in the final analysis. Two other *S*s failed to complete all of the experimental forms. A total of 50 *S*s comprised the experimental sample.

Equipment

The experiment was conducted in four small semi-soundproof rooms, one of which served as a master control room and the others served as experimental rooms in which the *S*s were isolated individually. It was possible, therefore, to have a maximum of three *S*s at one time. This number was optimal, since every effort was to have *S* believe that he was one of a group of three. Care was taken, however, to insure that *S*s did not see each other before the experiment, and it was possible to conduct the experiment even when only one *S* reported.

Each experimental room contained a desk and chair, and *S* was provided with a pencil, a set of headphones, and a microphone. The master control room contained four tape recorders. Separate recorders were available to record and/or monitor everything that happened in each of the three experimental rooms. The fourth recorder played a master tape that was heard over the headphones in all three of the experimental rooms.

Procedure

As soon as *S* was ushered into his experimental room, he was handed a form entitled Value Profile. This contained 20 statements which the *S* rated as to his agreement or disagreement on a seven-point scale. Ten of the items comprised the Self-Esteem scale which measures feelings of social inadequacy. These items were adapted from items used by Janis (1954). They had been previously shown to be highly intercorrelated and related to a conformity measure (Rosenbaum & deCharms, 1957). Ten items from the F scale were used as filler.

When *S* had finished the questionnaire, *E* gave instructions to each *S* individually concerning the remainder of the experiment, which was described as a study of the ability to evaluate others after a short nonvisual interaction. Since *E* visited each *S* individually, it was necessary for each *S* to sit quietly while the others were being instructed. The amount of time was roughly equated even in sessions when only one *S* was present.

The instructions prepared *S* to describe himself to two other participants following the topics and the order presented to him on the instruction sheet. This sheet made it clear to him when he was to speak and when the others were to speak. The sequence of speaking instructions included instances in which a single participant spoke to only one other person as well as occasions in which two others were addressed. The topics which each participant was to discuss were:

Round I. "Who are you?" This was to include relevant background information but no names.

Round II. Describe your social self.

Round III. Describe your nonsocial self.

Round IV. Characteristics that you dislike in yourself.

Round V. Characteristics that you like in yourself.

Round VI. Characteristics that you would like to possess ideally.

Brief explanations were given of each of these topics to give *S* a start. The instructions did not lead *S* to expect any spontaneous interaction with the other *S*s, since each was to be talking about himself and was not to ask or answer questions. Each *S* was told that he would be called B for the duration of the experiment. The people to whom he would be talking would be A and C. The order in which each spoke varied from round to round but the over-all order was constant for all conditions up to and including Round V. The order of speaking in Round VI varied according to the experimental condition.

TABLE 1

Sequence of Interactions in Round VI

Condition	Minute 1	Minute 2 (The Attack)	Minute 3
No Communication	B to A & C	A to B & C	C to A
Vicarious Communication	B to A & C	A to B & C	C to A & B
Communication	C to A	A to B & C	B to A & C

A master tape which each *S* heard contained instructions from the *E* explaining the topic to be discussed and telling who was to speak to whom. The tape also contained standardized prerecorded speeches for both A and C for each of the rounds.[2] When B (the *S*) was speaking, the tape was silent and the other tape recorders recorded what each *S* actually said about himself on each round. Three versions of the master tape were used, one for each communication condition, and these were identical up to and including Round V. On Round VI, A, who had up to this time presented himself as a self-assured, rather egotistical person, attacked B (the *S*), calling him, among other things, an "egotistical, deceitful liar" and alleging that he was "full of bullshit." These phrases were taken from Thibaut and Coules (1952). This attack lasted for one minute as did all the other communication periods for each person within a single round. The attack always came in the middle of Round VI, i.e., A always spoke second on this round and spoke to both B and C (see Table 1). He directed his remarks completely to B, however. The experimental manipulation in the different versions of the tape on Round VI are presented in Table 1. In the No-Com condition, B spoke first to A, then A spoke to B and C (the attack on B), then B was not allowed to respond but had to sit in silence (as a time control) while C was supposed to be talking to A. In the Com condition, C spoke first to A while B sat in silence. Then A spoke to B and C (the attack), and then B was allowed to respond to A. In the Vi-Com condition B spoke first to A, then A to B and C (the attack) and then C attacked A (while B listened) for throwing his weight around, calling him "fantastic" and a "bastard" who "ought to live in a jungle." Care was taken that C did not mention B at all in his attack on A.

Immediately after Round VI, *E* visited each *S* individually, giving him a form on which he was to write a first draft of a personality sketch of both A and C, in that order, in which *S* was asked (*a*) to "Describe his personal characteristics as you see them," (*b*) "What characteristics of his do you like and/or dislike?" (*c*) "Under what circumstances, if any, and how frequently would you like to associate with this person?" (*d*) "Please give your over-all evaluation of him."

In order to increase the likelihood that *S* would perceive the personality sketch as a means for aggressing against the attacker, he was told that the person for whom the form was filled out would be the only person to see it and that he would be allowed to make comments on it and return it to the *S*, at which time *S* would fill out the final draft to be turned in to *E*. In order to facilitate this procedure, *S*s were provided with envelopes addressed to A and C, and they enclosed their "first drafts" in the envelopes which were "delivered" by *E*.

---

[2] We wish to thank Robert Gestaldo and Robert Morris for playing the roles on the master tape.

When *S* had completed the personality sketch, the experimental session was essentially over. However, *E* visited each *S* individually, telling him that he had finished first and would have to wait to receive the ratings from the others. An interview was then carried out, starting with an inquiry about the electrical equipment and leading up to *S*s feelings about the situation. *E* attempted in every case (although he was not always successful) to get *S* to tell him about the attack, which *E* was not supposed to have heard. Once the attack was mentioned, *S* was encouraged to describe his feelings and was further pushed extensively to determine whether he had been at all suspicious of the situation. Once this information was obtained, the *S*s were brought together and, after the deception had been revealed, they were again asked about suspicion. The entire experiment was then explained in detail and rapport reestablished. Before leaving the experimental rooms *S*s were enjoined not to talk about the experiment. No *S* reported having heard anything about the experiment prior to appearing.

Analysis of the Data

The self-esteem score for each individual was the sum of his responses on a 7-point scale to the 10 self-esteem items on the pre-experimental value profile. Since half of these items were worded negatively, contamination by response set was avoided. High and low self-esteem scores were determined by a median split of all *S*s in the experiment. The distribution of scores was not skewed, and there were no significant differences between mean self-esteem scores in the three experimental treatments.

The primary dependent variable in the experiment was a residual hostility score derived from a content analysis of the "personality sketch." The responses on this form were first separated into thought units, usually phrases, and then each phrase was categorized as (*a*) positive evaluation of the person, (*b*) neutral, or (*c*) negative evaluation. Scoring was done by two scorers who knew nothing about the experiment.[3] Interscorer reliability was 92% agreement.

The score used as the measure of residual hostility was the percentage of thought units categorized as negative. Since there were very few neutral units and they were equally distributed, a score based on the positive units resulted in the same conclusions. There were no significant differences in mean total and number of units produced between experimental groups. "Residual hostility score" refers henceforth to the percentage of thought units categorized as negative in the personality sketch of the aggressor, A.

TABLE 2

MEAN RESIDUAL HOSTILITY SCORES (% NEGATIVE UNITS)

	No Communication	Vicarious Communication	Communication	Total
High self-esteem				
Mean	49.73	47.28	53.14	50.00
N	11	7	7	25
Low self-esteem				
Mean	68.28	49.60	42.63	52.60
N	7	10	8	25
Total				
Mean	56.94	48.65	47.53	
N	18	17	15	

RESULTS

Table 2 presents the residual hostility means by experimental conditions. The total means of the three communication conditions are in the order predicted by the hypotheses; namely, residual hostility is greatest in the No-Com condition, least in the Com condition, and intermediate in the Vi-Com condition. However, analysis of variance demonstrates that these differences are not significant. A test of the difference between the Com and No-Com conditions (Hypothesis I) reaches the .05 level of significance, confirming Thibaut and Coules' finding, although the test can be questioned since the over-all test does not reach significance.

Inspection of the body of Table 2 and of Fig. 1 makes it clear, however, that these tests, which ignore the level of self-esteem, obscure real differences. The difference apparent in Fig. 1 (solid lines) between the pattern of the high self-esteem means and that of the low self-esteem means in the evaluations of A is statistically significant.[4] It is obvious, therefore, that Hypothesis III is supported in the sense that the predicted effects occur most strongly in the low self-esteem groups. However, implicit in the hypothesis was the prediction that the effects would be similar but of a lesser magnitude for high self-esteem groups. This is clearly not supported. The high

---

[3] Ruthe Harriger and Marvin F. Daley.

[4] Both a test of interaction within an analysis of variance of the data in Table 2 and a test of the difference between the slopes presented in Fig. 1 reach statistical significance. The estimate of error derived from the analysis of variance was employed in further statistical tests when appropriate.

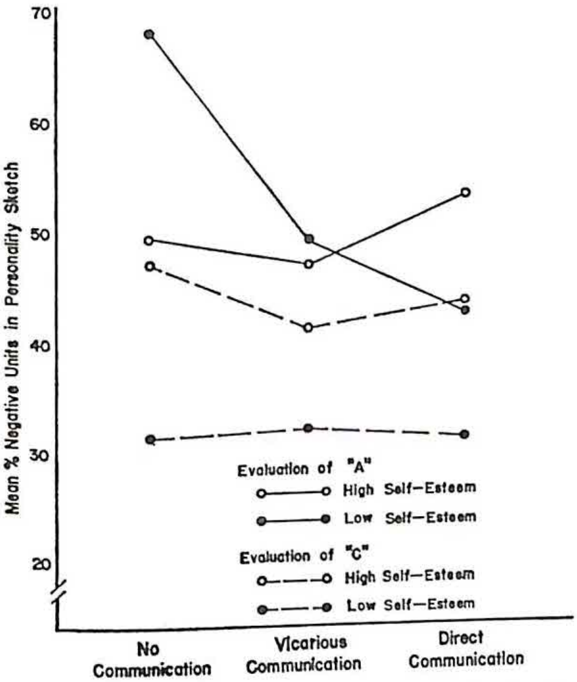


FIG. 1. Mean percentage of negative units in the personality sketches of "A" and "C."

self-esteem means are not in the order predicted and do not differ significantly. However, an analysis of variance of the three low self-esteem means demonstrates a significant difference ($p < .05$).

A more detailed analysis employing one-tailed t tests gives the following results:

Hypothesis I predicts a lower mean (less residual hostility) for the Com condition than for the No-Com condition. This hypothesis is supported for the *S*s with low self-esteem ($t = 2.61$, $df = 44$, $p < .01$).

Hypothesis II predicts a lower mean (less residual hostility) for the Vi-Com condition than for the No-Com condition. This hypothesis is supported for *S*s with low self-esteem ($t = 1.91$, $df = 44$, $p < .035$).

No significant difference can be demonstrated between the Vi-Com and the Com conditions. It is, therefore, impossible to show conclusively that the Com condition reduces hostility more efficiently than the Vi-Com condition, although the directional trend suggests that such is the case.

A subsidiary analysis of interest involves *S*'s evaluation of C. One might expect C to be more positively evaluated in the Vi-Com condition. In this situation, in contrast to the other two conditions, C behaved in a manner that was clearly negative toward A, although C was not at any time directly supporting of B (*S*). This should result in a more positive evaluation of C, especially for *S*s with low self-esteem, since it has been demonstrated that their responses to A were affected by the Vi-Com situation.

Content analysis of the personality sketches for C results in no support for this expectation (see Fig. 1, dotted lines). No statistically significant differences in evaluation of C can be demonstrated between the various communication conditions for either high or low self-esteem. There is, however, a significant difference ($p < .025$) between high and low self-esteem *S*s over all communication conditions, indicating that the low self-esteem *S*s (bottom line, Fig. 1) rate C more positively than do the high self-esteem *S*s in all conditions.

DISCUSSION

Hypotheses I and II predicted a greater amount of residual hostility in the No-Com condition than in either the Vi-Com or the Com conditions regardless of level of self-esteem. Hypothesis III predicted a difference between high and low self-esteem *S*s, but primarily in terms of relative differences in instigation to aggression. Since less instigation was expected for high self-esteem *S*s, it was expected that Hypotheses I and II would hold for these *S*s, but that the relationships found would not be as marked. The results support Hypotheses I and II unequivocally for *S*s with low self-esteem. Hypothesis III is supported in the sense that the results predicted by the first two hypotheses were indeed found most strongly with low self-esteem *S*s. No significant relationships were found, however, with high self-esteem *S*s. The failure of the predicted relationships to appear in the results for *S*s with high self-esteem must be accounted for.

Several possibilities for an explanation of the behavior of the high self-esteem *S*s are available. First, it is possible that these *S*s persist in hostile responses regardless of the means of expression available to them; i.e., after instigation to aggression no type of response opportunity presented in the experiment affects the level of residual hostility.

A counterproposal is that the high self-esteem *S*s were not instigated to aggression, and the amount of residual hostility manifested is not related to the insult procedure. The evidence to follow would seem to favor the latter alternative. The assumption can be offered that residual hostility in the No-Com condition reflects a base line for amount of hostility, in that neither personal nor vicarious expression changes the level of residual hostility. Within the No-Com condition regression analysis indicates a significant difference ($p < .01$, 2 tails) between the high and low self-esteem *S*s, the latter manifesting more residual hostility. This finding, which can be seen in Fig. 1, tends to indicate a lower level of instigation for the high self-esteem *S*s, as was expected when the need for maintenance of self-esteem is the activity interfered with.

Since instigation was apparently low for *S*s with high self-esteem, it is possible that they were not sufficiently instigated for aggressive responses to play an important role in the subsequently required personality sketches. This possibility would, of course, lead to lack of differences in residual hostility as a function of the communication conditions.

Some evidence can be brought to bear to support this possibility. Although no difference in evaluation of C in the different conditions was found, when a comparison is made of the difference in evaluation of A and C the results indicate that low self-esteem *S*s present a significantly lower mean percentage of negative thought units in evaluating C than in evaluating A ($t = 3.50$, $p < .01$, 2 tails). The difference for the high self-esteem *S*s although in the same direction is not significant ($t = .97$, $p > .30$, 2 tails).

It appears that the high self-esteem *S*s in general perceived no difference between A and C, despite the fact that A under all conditions attacked them while C did not. This finding would seem to indicate a certain insensitivity to the comments of others on the part of high self-esteem *S*s. The attacking statements made by A would appear to have failed to arouse anger in the high self-esteem *S*s. This manifestation of insensitivity is consistent with other findings associated with the self-esteem variable. High self-esteem *S*s have been found to be relatively nonresponsive to persuasion (Janis, 1954) and to efforts to elicit conformity (Rosenbaum & deCharms, 1957). Persons with high self-esteem do not appear to respond readily to stimuli presented by the behavior of others.

The results of this study represent a replication and clarification of the findings of Thibaut and Coules (1952). Some degree of reduction of the equivocality of their results may be claimed, although, as will be discussed, this reduction may not be complete. In contrast to the procedure of the earlier study, the *S*s in the No-Com condition of the present experiment were not summarily interrupted without warning. They were informed at the beginning of the interaction of the sequence which was to end without a rebuttal opportunity. They nevertheless may have been considerably frustrated, an effect which may have augmented the aggression which was expressed within the personality sketch. Some doubt is created for this explanation, which was presented by Thibaut and Coules, in the finding of the present experiment that the Vi-Com *S*s, who also were frustrated by lack of opportunity for self-expression, nevertheless present significantly less residual hostility than the No-Com *S*s. It is, of course, possible that the augmented frustration still operates for the No-Com group and the Vi-Com situation serves successfully to relieve the frustration which accrued from lack of own expression. No way of deciding between these two alternatives is available in this experiment.

The catharsis hypothesis, which was the starting point for the Thibaut and Coules study and for the present experiment, apparently receives support from the results. It must be pointed out, however, that no direct evidence is available from either study to indicate the mechanism responsible for the lowered residual hostility in the Com condition as compared to the No-Com. In the present experiment there is scant evidence that aggressive verbalizations during the rebuttal are the cause of the reduced hostility in the Com condition. The rebuttal statements of the *S*s were far milder than would be expected if a response in kind was forthcoming. That is to say, the general rebuttal response appeared far less aggressive than the standardized attack delivered by A. For example, although A employed profanity in his attack, none of the Com *S*s responded in this manner. In fact, clearly aggressive words were extremely rare and never approached the level of those used by A. Very frequently Com *S*s said nothing relevant to the attack, but proceeded in Round VI to carry out the task of describing their "ideal self." In general, it seems unlikely that the verbalizations were of a character to lead to a reduction of hostility.

These results strongly suggest that a catharsis hypothesis, i.e., reduction of hostility due to expressed aggression, may be inadequate in accounting for the findings of this and the Thibaut and Coules experiment. Clarification of the mechanism which accounts for the hostility reduction implicit in this study awaits further research.

SUMMARY

Fifty *S*s participated in a simulated interaction in which each was verbally attacked by a standard tape recording. Under three experimental conditions *S*s (*a*) sat in silence after the attack, (*b*) heard a third person denounce the attacker, (*c*) were permitted to communicate back to the attacker themselves. *S*s in each group were classified as to level of self-esteem. A measure of residual hostility toward the attacker constituted the primary dependent variable.

Results demonstrated (*a*) more residual hostility for low self-esteem *S*s who were not allowed to reply to the aggressor than for similar *S*s who could communicate or who heard someone else denounce the aggressor, (*b*) no significant differences between groups of high self-esteem *S*s, (*c*) differences between the evaluation of the aggressor and of the third person for low self-esteem *S*s, but not for high self-esteem *S*s, and (*d*) very little actual aggression shown by any of the *S*s who were allowed to communicate back to the aggressor after the attack.

These results give support to the hypotheses that both direct and vicarious communication to the aggressor reduces hostility for *S*s with low self-esteem. It appears that high self-esteem *S*s were not very strongly instigated to aggression, and that they showed some insensitivity to cues presented by other persons since they seemed not to discriminate between the aggressor and the other more benign person. An explanation of the results of this and of Thibaut and Coules (1952) study using a catharsis hypothesis was called into question, since very little actual aggression was observed in *S*s response to the agressor, and yet this communication, which was in most instances unaggressive and unrelated to the attack, appears to have reduced hostility.

REFERENCES

BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, **55**, 257–283.

DECHARMS, R., & ROSENBAUM, M. E. The problem of vicarious experience. In Dorothy Willner (Ed.), *Decisions, values, and groups.* New York: Pergamon, in press.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3–11.

FRENCH, J. R. P. Organized and unorganized groups under fear and frustration. *Univ. Ia. Stud. Child Welf.*, 1944, **20**, 229–308.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.

PEPITONE, A., & REICHLING, G. Group cohesiveness and the expression of hostility. *Hum. Relat.*, 1955, **3**, 327–339.

ROSENBAUM, M. E., & DECHARMS, R. Status revocability and social conformity. *USN ann. Tech. Rep.*, Nov. 1957, ONR855(04), Org. Res. Group, Univer. of North Carolina. Pp. 6–7.

THIBAUT, J. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, **3**, 251–278.

THIBAUT, J. W., & COULES, J. The role of communication in the reduction of interpersonal hostility. *J. abnorm. soc. Psychol.*, 1952, **47**, 770–777.

(Received January 19, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

PREDICTION OF SMALL GROUP ROLE BEHAVIOR FROM PERSONALITY VARIABLES

WALTER R. BORG
Utah State University

THIS study explores the degree to which a person's role in a small group problem solving seminar can be predicted from a battery of group administered tests which attempt to measure skills and personality characteristics observed to be important in small group situations. The specific hypothesis to be tested is that a person's scores on tests selected to give measures of assertiveness, rigidity, self orientation, and sociability will predict his role in a problem solving seminar.

METHOD

Subjects

The *S*s included 819 Air Force Majors and Lieutenant colonels in the Command and Staff School of the Air University, Class 1957. They were divided into 60 seminars of 12 to 14 officers each, and worked in the same seminar group for a period of two and one-half months. A large portion of the curriculum of the school is devoted to problem solving activities and discussions in the seminar groups. The problems dealt with by the seminars vary considerably in nature but, for the most part, involve practical Air Force problems of a sort that might be encountered by a Wing Staff.

Variables and Measures

The first step in the research was the selection of tests to go into the predictor battery, which was to contain tests to measure variables related to patterns of behavior that had been found in observations of small group activity. A careful survey was made in order to identify variables that had occurred repeatedly in studies based on observations of small group interaction (Bass, 1951; Bass, McGehee, Hawkins, Young, & Gebel, 1953; Carter 1951, 1954; Carter, Haythorn, & Howell, 1950; Carter, Haythorn, Shriver, & Lanzetta, 1951; Cattell & Stice, 1953; Haythorn, 1953; Sakoda, 1952; Schutz, 1955). The variables that appeared most frequently in these studies were:

Sociability. This variable, under different names, such as Personal Orientation, Group Sociability, and Social Adjustment, has been reported frequently in studies concerned with small group interaction.

Self orientation. Variations of this measure have appeared repeatedly under such names as Power Orientation (Schutz, 1955) and Individual Prominence, (Carter, 1954; Carter & Nixon, 1949).

Rigidity. Rigidity and related variables such as Authoritarianism, Conformity, and Dogmatism have been investigated in a number of relevant studies of small group behavior (Greer, 1955; Haythorn, Couch, Haefner, Langham, & Carter, 1956; Lippitt, 1940; McCurdy & Eber, 1954; Medalia, 1955).

Assertiveness. This variable has been employed with considerable success by Schutz (1955) and is somewhat similar to variables employed in other small group research such as Forceful Leadership, Surgency (Cattell, 1951), Ascendance (Bass et al., 1953), Physical Energy (Sakoda, 1952), and Individual Prominence (Carter, 1954).

The tests selected to supply predictive measures of these characteristics included the following:

The FIRO test. This test was developed by Schutz (1955) and has been used by him with considerable success for predicting leadership behavior and organizing compatible groups. The form of the test employed in this research consisted of seven items, each of which was regarded as a separate score on the suggestion of Schutz. These items attempt to measure degree of power orientation, as compared with personal orientation and degree of assertiveness.

The California F Scale, Form 40. This form of the F scale employs the items that were found to be most discriminating in the original research at the University of California (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950).

The Anxiety to Achieve Battery. This battery consisted of six group tests from the Cattell Objective Analytic Personality Battery (Cattell, 1955) and yielded a single factor score.

The Social Extroversion Score, from the Inventory of Factors STDCR (Guilford, 1945).

Eight factor scores from the Guilford Opinion Inventory (Guilford, 1951). This test was developed by Guilford from his factor analysis of the "interests" of Air Force officers. Factors used included: need for variety, orderliness, need for attention, physical drive, aggression, resistance to restriction, cultural conformity, self-reliance. The entire predictor battery yielded 18 scores.[1]

Derivation of Factor Scores

After the battery was administered, results were subjected to centroid factor analysis. A total of four orthogonal rotations were made, yielding four factors.

---

[1] Description data on the predictor battery and the six role descriptions used in the peer nominations have been deposited with the American Documentation Institute. Order Document No. 6094, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
ROTATED FACTOR LOADINGS EMPLOYED IN DEVELOPING WEIGHTED FACTOR SCORES

Variable	Loading
Factor 1: Assertiveness	
Schutz G-H, Assertiveness	.61
Schutz 5, Assertiveness	.66
Schutz 6, Success in Attaining Leadership	.57
Schutz 7, High Self-Estimate of Leadership Ability	.43
S Factor from Guilford's STDCR, Social Extroversion	.58
P Factor from Guilford's Opinion Inventory, High Level of Physical Drive	.36
Factor 2: Power Orientation	
Schutz A-B, Power Orientation	.43
G Factor from Guilford's Opinion Inventory, Self-Reliance	.40
Factor 3: Rigidity	
California F Scale, Authoritarianism	.56
S_0 Factor from Guilford's Opinion Inventory, Tendency toward Orderliness	.39
F Factor from Guilford's Opinion Inventory, Cultural Conformity	.53
Factor 4: Aggressive Nonconformity	
V_0 Factor from Guilford's Opinion Inventory, Need for Variety	.45
L Factor from Guildord's Opinion Inventory, Need for Attention	.54
Q Factor from Guilford's Opinion Inventory, Aggression	.62
M Factor from Guilford's Opinion Inventory, Resistance to Restriction	.43

Rotated factor loadings may be found in Table 1. The factors were interpreted as follows:

Factor 1, Assertiveness, was loaded heavily on Schutz's Assertiveness scores, as well as Guilford's Physical Drive score and Social Extraversion score. A separate Sociability Factor, mentioned as an important variable in several studies, did not emerge. If additional Sociability measures had been included in the predictor battery, the factor may have emerged or may have loaded under Assertiveness as did Guilford's Social Extraversion score.

Factor 2, Power Orientation, had loadings on Schutz's Power Orientation score, as well as Guilford's Self-Reliance variable.

Factor 3, Rigidity, seems clearly indicative of Rigidity, being loaded most heavily on the California F scale. Other significant loadings include Guilford's Cultural Conformity score and Orderliness score.

Factor 4, Aggressive Nonconformity, is more difficult to interpret. The heaviest loading on this factor was in Guilford's Aggression scores. Other variables loading on this factor included Guilford's Need for Variety score, Need for Attention score, and Resistance to Restriction score. Several interpretations of this factor were considered. In some respects it appeared quite similar to Cattell's Hypomanic Aggression. Another possible interpretation was that the factor was indicative of social immaturity, while a third interpretation was that it reflected frustration in the military environment. For several reasons, an interpretation of the factor as Aggressive Nonconformity seemed most plausible. First, the heavy weighting of Aggression and Need for Attention seemed to point in this direction. The loadings on Need for Variety and Resistance to Restriction suggest a negative reaction to the restrictions and conformity demands found in the military environment. Observations of the Command and Staff School, secondly, reveal that a significant number of students display overt aggressive attitudes toward the school. Some students also displayed considerable negativism concerning the testing itself. This negativism might well have been reflected in some of the test scores where it could have come to the surface. For example, some of the items on the Aggression test, such as those advocating punitive action, could give the *S* an indirect means of expressing aggression or negativism aroused by the testing situation. Weiss' (1956) research dealing with the effect of the individual's immediate mood on attitude measures indicates that the individual's mood may be reflected to a significant degree in items where aggression can be displayed.

Weighted factor scores were computed including all variables loading over .30 and most heavily loaded on the given factor. These personality factor scores constituted the predictors against which to compare small group role scores (see Table 1).

Small Group Role Scores

The small group role scores were based on two peer nominations completed in the second and fifth weeks after the groups were organized. The first peer nomination contained six brief paragraphs describing typical small group behavior patterns. These paragraphs described popular-social behavior, good followers, assertive behavior, rigid behavior, creative behavior, and leadership. Each seminar member nominated the person (by code number) who best fitted each description and checked a five-point scale to indicate the degree to which the selected individual fitted the description. Each seminar member received a score for each of the six roles which was the sum his nominations weighted for degree of fit. These scores were then converted to standard scores.

The second peer nomination used the same six role descriptions but required the individual to select three seminar members most fitting the description and three least fitting the description. The individuals' score for each role was the algebraic sum of his best fit and least fit nominations. Scores on the second peer nomination were converted to standard scores and combined with first peer nomination scores to give six composite role scores for each *S*.

RESULTS

In order to determine the degree to which the personality factor scores predicted small group role scores, samples of individuals who had made either high or low role scores were selected. The four individuals in each seminar who received the highest scores on a given role were selected as well as the four indi-

TABLE 2

CORRELATIONS BETWEEN PREDICTOR FACTORS AND SMALL GROUP ROLE SCORES

	Popular-Social	Good Follower	Assertive	Rigid	Creative	Leader
Factor Scores....................	1	2	3	4	5	6
1—Assertive	.245*	−.221*[a]	.422*[a]	.216*[a]	.423*[a]	.389*[a]
2—Power Orientation	−.044	.066	.012[a]	.008[a]	.020[a]	.020
3—Rigidity	.065[a]	.026[a]	−.114*	.012	−.176*[a]	−.158*[a]
4—Aggressive Nonconformity	.108	−.060	.141*[a]	−.020	.207*[a]	.175

*Significant beyond .01 level.

[a] Results as hypothesized.

viduals with the lowest role scores. These cases were further screened to eliminate individuals from the top group who were not at least .6 σ above the mean on both peer nominations and eliminating individuals from the bottom group who were not at least .6 σ below the mean on both peer nominations. Thus, the persons chosen were not only at the extreme within their group but were at the extreme with respect to the entire sample on the role score under consideration.

A total of 24 specific hypotheses were then developed. The relationship of each of the four predictor factors to each of the six small group roles was hypothesized to be either positive, negative, or zero (see Table 2) based upon psychological theory and previous research results. Fourteen of these hypotheses were supported by the data. In all, 12 of the 24 widespread biserial correlations computed between the small group roles and the personality factors were significant at the .01 level.

It may be seen, however, in Table 2 that although many of the correlations were significant they were too small to permit prediction of the role of a given individual from his personality scores. By far the most promising of the predictor scores is Factor 1, Assertiveness. This score correlated significantly with all six role scores and yielded correlations around .40 with Assertiveness, Creativity, and Leadership roles. As Table 3 indicates, these three roles are moderately intercorrelated, suggesting that they are measuring a single broad leadership role. A combination of these three role scores correlates .46 with Factor 1. This correlation is sufficiently high to permit reasonably good selection for small group leadership if it is established that this role is stable from situation to situation.

TABLE 3

CORRELATIONS BETWEEN DIFFERENT PEER ROLES ON THE SECOND RATING

Variable	1	2	3	4	5	6
1. Popular-social	.000	.139	.255	−.348	.264	.252
2. Good Follower	.139	.000	−.251	−.404	.065	.105
3. Assertive	.255	−.251	.000	.452	.710	.599
4. Rigid	−.348	−.404	.452	.000	.207	.177
5. Creative	.264	.065	.710	.207	.000	.817
6. Leader	.252	.105	.599	.177	.817	.000

DISCUSSION

The results shown in Table 2 generally support the findings of earlier research. Several studies have been reported that are concerned with relationships between personality and small group behavior. Some of these have analyzed small group observational data so as to yield personality factors. Some have explored relationships between peer or supervisor evaluations and small group behavior. Only a few, however, have attempted to predict small group behavior from personality test scores.

Leadership ratings of 20 sorority girls in a leaderless group discussion situation were compared with their scores on the Rorschach, Guilford-Zimmerman Temperament Survey and F scale by Bass (1953) in one study of this type. Bass found no significant results with the F scale, but did find that Ascendence and Sociability scores on the Guilford-Zimmerman and High Verbal Output on the Rorschach were significantly correlated with Leadership ratings. The Rorschach result is in all likelihood an artifact of Bass' LGD rating system, which depends to a considerable degree on Verbal Output (Bass, 1951). The correlation between Ascendency and Leadership in Bass's study, however, is almost identical to the correlation between Assertiveness and Leadership found in the research reported in this paper.

In another study of small group leadership (Cattell & Stice, 1953) leaders in 34 groups were compared in terms of scores on the 16 PF test. He found four variables that differentiated between leaders and nonleaders at the .05 level. These were Super Ego Strength, Adventurousness, Self Confidence, and High Self-Sentiment Formation. These variables, as described by Cattell, are similar in many respects to the Assertiveness and Power Orientation factors used as predictors in this paper. Richardson and Hanawalt (1944) compared two types of leaders and nonleaders on the basis of their scores on the Bernreuter Personality Inventory. They found their leader groups to be more dominant, extraverted, and self-sufficient than the nonleaders.

Thus, the few studies that have been done, despite the use of different personality measures, different leadership measures, and different types of *S*s, all agree to a considerable extent that leaders are significantly more assertive, self confident, and extraverted than nonleaders.

Studies employing other personality variables as predictors of small group roles are much less conclusive.

Several workers have employed the F scale and related measures in small group research with varying results. The rigidity factor used in the research reported in this paper was correlated negatively to the variables in the broad leadership role, but the correlations, although significant, are too low to have much meaning. Some work has indicated that authoritarian leaders function effectively in authoritarian groups (McCurdy & Eber, 1954). There is also evidence that different types of individuals emerge as leaders in authoritarian and equalitarian groups (Haythorn et al., 1956; Adorno et al., 1950). Some work has demonstrated a positive relationship between Authoritarian Leadership and Group Cohesiveness (Medalia, 1955), while other research has found a negative relationship between these variables (Lippitt, 1940). Unlike Assertiveness, which appears to be an important leadership variable in most groups, the importance of authoritarianism varies greatly with the group, the task, and the leader.

Four of Schutz's items loaded heavily in the Assertiveness factor and one in the Power Orientation factor identified in this research. The results of Schutz's experimental battery, particularly the Assertiveness items, were impressive but indicated a need for further development. Schutz's scores have rather low reliability as they are each based on only one item. The items aimed at measuring Assertiveness were found to correlate with each other from .19 to .47. Items measuring Power vs. Personal Orientation, however, correlated with each other $-.09$ to $+.09$, indicating little common variance. As Schutz's technique becomes more highly developed, the Power Orientation variable might become useful in role prediction.

SUMMARY AND CONCLUSIONS

A sample of 819 Air Force officers was administered a test battery designed to predict the individual's role in small group situations. This battery was factor analyzed, yielding four factors: Assertiveness, Power Orientation, Rigidity, and Aggressive Nonconformity. The sample was then divided into 60 small groups of 12 to 14 members and rated with respect to six small group roles. Twenty-four widespread biserial correlations were computed between the four predictor factors and the six small group roles. Twelve of these correlations were significant at the .01 level. The predictor factor "Assertiveness" was most successful, correlating .46 with a composite leadership role.

With regard to significance of these results for future work, it seems reasonable to conclude from the success in predicting the leadership composite that prediction of certain roles and behavior patterns in small group activity can be achieved by further developing predictor instruments along the lines indicated by this study. Further research may make it possible to analyze changes in the individual's role behavior that are attributable to differences in certain measurable characteristics of other group members in problem solving groups.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BASS, B. M. Situational Test: 2. Leaderless group discussion variables. *Educ. psychol. Measmt.*, 1951, **2**, 196–207.

BASS, B. M., MCGEHEE, C. R., HAWKINS, W. C., YOUNG, P. C., & GEBEL, A. S. Personality variables related to leaderless group discussion behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 120–128.

CARTER, L. F. Some research on leadership in small groups. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951.

CARTER, L. F. Evaluating the performance of individuals as members of small groups. *Personal Psychol.*, 1954, **7**, 477–484.

CARTER, L. F., HAYTHORN, W., & HOWELL, M. A further investigation of the criteria of leadership. *J. abnorm. soc. Psychol.*, 1950, **45**, 350–358.

CARTER, L. F., & NIXON, M. An investigation of the relationship between four criteria of leadership for three different tasks. *J. Psychol.*, 1949, **27**, 245–261.

CARTER, L. F., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.

CATTELL, R. B. *Handbook for the objective-analytic personality test batteries*, Institute for Personality and Ability Testing, Champaign, Illinois, 1955.

CATTELL, R. B. Determining syntality dimensions as a basis for morale and leadership measurement. In H. Guetzkow (Ed.), *Groups, leadership and men*, Pittsburgh: Carnegie Press, 1951.

CATTELL, R. B., & STICE, G. S. *Four formulae for selecting leaders on the basis of personality.* Urbana, Ill. Univ. of Illinois, Lab. Pers. Assessment & Group Behav., 1953.

GREER, F. L. *Small group effectiveness, Series 1955.* Philadelphia: Inst. for Res. in Hum. Relat., 1955.

GUILFORD, J. P. *Opinion Inventory.* San Antonio, Tex.: Air Res. and Devel. Command, 1951.

GUILFORD, J. P. *Manual. An inventory of factors STDCR.* Beverly Hills, Calif. Sheridan Supply, 1945.

HAYTHORN, W. The influence of individual members on the characteristics of small groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 265–284.

HAYTHORN, W., COUCH, A., HAEFNER, D., LANGHAM, P., & CARTER, L. The behavior of authoritarian and equalitarian personalities in groups. *Hum. Relat.*, 1956, **9**, 57–74.

LEWIN, K., LIPPITT, R., & WHITE, R. Patterns of aggressive behavior in experimentally created social "climates." *J. soc. Psychol.*, 1939, **10**, 271–299.

LIPPITT, R. An experimental study of authoritarian democratic group atmospheres. *Univ. Ia. Stud. Child Welf.*, 1940, **16**, 43–195.

MCCURDY, H. E., & EBER, H. W. Democratic vs. authoritarian: A further investigation of group problem solving. *J. Pers.*, 1954, **22**, 258–269.

MEDALIA, N. Z. Authoritarianism, leader acceptance, and group cohesion. *J. abnorm. soc. Psychol.*, 1955, **51**, 207–213.

RICHARDSON, HELEN M., & HANAWALT, N. G. Leadership as related to the Bernreuter Personality Measures: III. Leadership among adult men in vocational and social activities. *J. appl. Psychol.*, 1944, **28**, 308–317.

SAKODA, J. M. Factor analysis of OSS situational tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 843–852.

SCHUTZ, W. C. What makes groups productive? *Hum. Relat.*, 1955, **8**, 429–465.

WEISS, W., & FINE, B. J. The effect of induced aggressiveness on opinion change. *J. abnorm. soc. Psychol.*, 1956, **52**, 109–114.

(Received January 19, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

CHARACTERISTICS AND CONCOMITANTS OF MOOD FLUCTUATION IN COLLEGE WOMEN

ALDEN E. WESSMAN, DAVID F. RICKS[1]
Harvard University
AND MARY McILVAINE TYL
Radcliffe College

THE pathologically extreme moods that occur in neurotic depressions and in the affective psychoses have been described in much the same way for centuries (Burton, 1621; Bellak, 1952); but the causes of these alterations are still debated, and their relations to normal variations in hedonic level are still unclear. The present study is based on the general proposition that the difference between pathological and normal mood changes is essentially one of degree rather than of type, and that the phenomena which have been so thoroughly described as characteristic of extreme moods also characterize, in less pronounced forms, the mood variations of normal individuals. This proposition is examined using three sets of data provided by a group of normal college women: (*a*) their self and ideal representations reported in relatively elated and relatively depressed states, (*b*) the direction of punitive responses and persistence of effort revealed by a projective frustration test taken in these states, and (*c*) the average level, the frequency, and the intensity of their mood fluctuations reported daily over a period of six weeks.

Few studies of hedonic level and affects in normal persons have been concerned with repeated measurements over a period of time. Flügel (1925), regarding previous studies as segmental, artificial, and of insufficient duration, reported an investigation of the temporal patterns and nature of affective processes of nine subjects who kept systematic and detailed introspective records of their feelings of "pleasure" and "unpleasure" and who made frequent self ratings on a seven-point hedonic scale for a minimum period of 30 days. He reported that the data for every subject showed a distinct preponderance of "pleasure" over "unpleasure," and that the duration of intense affective states (whether "pleasure" or "unpleasure") was considerably shorter than that of less intense feelings. He also found "evidences to suggest that those who tended to experience the most extreme degrees of feeling are on the whole less happy than those whose feelings are usually less intense."

Johnson (1937) asked 30 female college students to furnish daily self ratings on mood for 65 to 90 consecutive days and to serve in two experimental sessions, one in an elated and one in a depressed mood. She reported that the individual temporal patterns of daily mood showed varying degrees of rhythmic tendency. She found no apparent relation with menstrual periodicity. A fairly consistent feature of the temporal patterns was the longer duration of the elated as compared to the depressed moods. For individual subjects, measurable behavior differences were found in elated and depressed moods. The subjective correlates of depressed moods revealed, in lesser degree, the symptoms characteristic of affective psychopathology: reduced feelings of physical energy, felt loss of power and capacities, reduced outgoing friendliness, indecisiveness, diffidence, and withdrawal from social interaction.

The general proposition that differences between pathological and normal mood phenomena are primarily matters of degree rather than of type suggests a broad scale application of hypotheses derived from psychoanalytic investigations (Freud, 1917; Abraham: 1954a, 1954b; Rado, 1928; Fenichel, 1945) as well as the studies of normals reported above. The hypotheses below are of two varieties: first, hypotheses about general *intra-*

---

[1] This study was carried out by the senior authors in connection with research at the Psychological Clinic Annex, Laboratory of Social Relations, Harvard University, under Research Grants M-700 and M-1287 from the National Institute of Mental Health. (Project director: Henry A. Murray)

individual comparisons in elated and depressed states of (*a*) self and ideal conceptions, and (*b*) aggression and persistence of effort when frustrated; second, hypotheses relating *inter*-individual variations on the characteristic parameters of mood fluctuation to the relative stability and congruence of self and ideal conceptions.

Intra-individual Change in Depression and Elation

1. Self esteem, as measured by the correlation between the self and the ideal, is higher in elation than in depression.
2. The self concept changes more than the ideal concept with changes of mood.
3. The self concept is relatively favorable in elation, and relatively unfavorable in depression.
4. The ideal concept is relatively less "good" (and therefore less demanding) in elation, and relatively more "good" (and therefore more demanding) in depression.
5. There is greater internal (introceptive) preoccupation in depression than in elation, and less interest in the environment.
6. The frequency of intropunitive responses to frustration is higher in depression than in elation, and the frequency of extrapunitive responses is lower.
7. The frequencies of obstacle-dominated and ego-defensive responses to frustration are higher in depression, and the frequency of need-persistent responses is lower.

Concomitants of Interindividual Differences on the Parameters of Mood Fluctuation

8. Self consistency, as measured by the correlation between the self description in elation and self description in depression, is negatively related to (*a*) the relative intensity of mood fluctuations, and (*b*) the relative frequency of pronounced mood fluctuations.
9. Self esteem, as measured by the correlations between the self description and ideal description (in elation and depression) is positively related to average level of mood (i.e., general hedonic level).

METHOD

Subjects

*S*s were 25 Radcliffe College students, all from the same dormitory, who volunteered to take part in a six-week study (October–December 1957) that required 5–10 minutes every night for recording their feelings, together with taking certain psychological tests. Of the original 25 *S*s, 21 responded 30 days or more of the possible 42 days; their records were used in the final data analysis of the patterns of mood fluctuation. The tests of hypotheses, however, are based on the 14 *S*s who provided a sufficient record of their feelings and who also completed the psychological tests described below.

Record of Daily Moods

Daily self ratings were made for six weeks on 11 scales describing different affective states. Each *S* also indicated daily the state of her health, the felt pressure of immediate academic work, and whether or not she was menstruating. Only the results of the 11th scale, *Mood*, and the question concerning menstruation are considered in this report.

The *Mood* scale contained 10 phrases, ranging from expressions of extreme depression through more neutral feelings to those expressive of extreme elation. The phrases were chosen so that (*a*) there would be approximately equal subjective gradations between units, and (*b*) the ends of the scale would be so extreme that few *S*s would experience them, and these few would do so on rare occasions. This scale is given below:

10. Complete elation, rapturous joy, and soaring ecstasy.
9. Very elated and in very high spirits; tremendous delight and buoyancy.
8. Elated and in high spirits.
7. Feeling very good and cheerful.
6. Feeling pretty good, "OK"
5. Feeling a little bit low. Just so-so.
4. Spirits low and somewhat "blue."
3. Depressed and feeling very low. Definitely "blue."
2. Tremendously depressed. Feeling terrible, really miserable, "just awful."
1. Utter depression and gloom. Completely down. All is black and leaden. Wish it were all over.

The scale was marked each night just before retiring for: (*a*) the best *S* had felt during the day, (*b*) the worst *S* had felt during the day, and (*c*) *S*'s average for the day.

Assessment of Self Concept and Ideal Concept in Elation and Depression

A Q sort was constructed by asking each *S* to write 10 phrases which could describe her, a friend, or other girls at Radcliffe. Almost all of these descriptive phrases could be classified according to *area*, i.e., whether the phrase referred to (*a*) introceptive feelings, (*b*) interaction with peers, or (*c*) intellectual traits and work in college; and *evaluation*, i.e., whether it implied (*a*) favorable, (*b*) neutral, or (*c*) unfavorable judgment.

For the Q sort, 45 items were selected, 5 in each of the 9 categories of the design.[2]

When filling out the Q sort to describe self and ideal, the *S* was instructed to fit the 45 items to a forced distribution on a 7-point scale ranging from "-3, least characteristic," through "0," to "$+3$, most characteristic." The number of items to be fitted in the respective scale positions, -3 to $+3$, were 3, 5, 9, 11, 9, 5, 3.

This set of items was given to the *S*s 3 weeks after the recording of their daily moods had begun, with the instructions that they were to wait until two occasions, one in which their mood reached an extreme "high," and one an extreme "low," and at those times administer the tests to themselves. The "elated" and the "depressed" sets were given to the *S*s in sealed envelopes, and were identical, consisting of two sets of the same 45 items, the first (self concept) to be filled out "to give the most accurate picture of yourself as you really believe you are now," the second (ideal concept), to be done after the first and without comparing items, to be filled out "to give a picture of yourself as the kind of person you have hoped to become or have fancied yourself to be; the picture you describe should be the personal ideal for you." After the *S* filled in the sorts, she placed them in an envelope labeled with a pseudonym and gave them back to the experimenters. *S*s were requested not to show the sorts to, or discuss them with, each other.

Projective Reactions to Frustration in Depression and in Elation

The Rosenzweig Picture-Frustration Test, Standard Revised Form for Adults, was used to measure projective reactions to frustration in depression and in elation (Rosenzweig, Fleming, & Clarke, 1947). Following some preliminary exploration, the instructions were changed slightly to read: "Imagine what *you* would say if you were the other person, and write in the box the very *first* reply that comes into your mind. Please avoid the temptation to be overly humorous. *Remember: we want what you probably would say in such a situation.*"

Two Rosenzweig P-F tests were given to each *S* with the Q sort forms and were self-administered in the same way—one when the *S* felt very elated and one when she felt very depressed.

The responses were scored according to Rosenzweig's instructions (*a*) as to *direction of aggression*, either "extrapunitive," "intropunitive," or "impunitive"; and (*b*) as to *type of reaction*, either obstacle dominance, ego defense, or need persistence; and the percentage of the total responses for each *S* in each of the classifications was obtained. The protocols were scored and hypotheses tested independently by two of the investigators. While there were some minor differences in scoring, the results of the statistical tests of the hypotheses were identical.

RESULTS

Intra-individual Change in Depression and Elation

Hypothesis 1. Self esteem. Of the 14 *S*s completing this part of the experiment, 13 showed a higher correlation between the self and the ideal in elation than in depression ($p < .005$, sign test[3]). In depression the correlations ranged from .78 to $-.83$, with a median of .07; in elation they ranged from .87 to $-.14$, with a median of .55. The median difference between the elated self-ideal correlation and the depressed self-ideal correlation was .45. If correlation between self and ideal can be regarded as a measure of self esteem we can conclude that, for these *S*s, one consistent concomitant of a change from elation to depression is a drop in self esteem.

Hypothesis 2. Relative Consistency of the Self and Ideal. The correlation between the ideal in depression and the ideal in elation was higher than the corresponding correlations for the self concepts for all 14 *S*s ($p < .005$). The correlations between the two ideal sorts ranged from .93 to .72, with a median value of .84. The correlations between the self sorts ranged from .79 to .18, with a median value of .63. The median difference between the two correlations was .26. The data indicate that in mood swings, the ideal remains relatively more stable than the self concept.

Hypothesis 3. Favorableness of the Self Concept. For all 14 *S*s the score for "unfavorable" statements regarding the self was greater, i.e., more characteristic, in depression than in elation, as predicted ($p < .005$). It was expected that this relation would hold for all 3 subareas of the Q sort; the relation was confirmed for the areas of "introception" (13 of 14 *S*s, $p < .005$) and "interaction" (13 of 14 *S*s, $p < .005$), but it was not supported for the area of "work" (8 of 14 *S*s, not significant).

---

[2] The set of structured Q-sort items has been deposited with the American Documentation Institute. Order Document No. 6095, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] All of the probabilities reported in the section on intra-individual changes with mood are based on one-tailed sign tests (Mosteller & Bush, 1954, pp. 312–314).

For 13 of 14 *S*s the score for "favorable" items regarding the self was smaller, i.e., less characteristic, in depression than in elation ($p < .005$). The predicted relationship held for all three areas of the Q sort: "introception" (12 of 14 *S*s, $p < .025$), "interaction" (13 of 14 *S*s, $p < .005$), and "work" (12 of 14 *S*s, $p < .025$). The results bear out the hypothesis that depression is characterized by an increase in self-derogatory attitudes and a decrease in feelings of goodness and worth.

Hypothesis 4. Favorableness of the Ideal. It was predicted that the ideal concept, unlike the self concept, would appear more favorable, more "good," in depression than in elation (and thereby increase the distressing disparity between the ideally desired self and the presently perceived "real" self). This was not confirmed for either "positive" items (6 of 12 *S*s, 2 ties, not significant) or for "negative" items (4 of 10 *S*s, 4 ties, not significant).

Inspection of the three subareas in the Q sort showed that for "interaction" and for "work" there were no significant findings in either the predicted direction or its opposite. In the subarea of "introception," however, the ideal was significantly more "favorable" or "good" in depression (10 of 13 *S*s, 1 tie, $p < .05$). For negative items the data approaches significance (8 of 11 *S*s, 3 ties) in the predicted direction. We conclude that the ideal is relatively less affected by mood changes than is the self (Hypothesis 2); but that in the area of introceptive feelings and evaluations the ideal is more "good," and thereby more discrepant from the actual self, in depression than in elation.

Hypothesis 5. Degree of Introception. This hypothesis suggested a greater preoccupation with "introceptive" judgments in depression, with less emphasis upon the persons and pursuits involved in "interaction" and "work" items. The items were accordingly scored, not only for the degree to which they characterized the self, but also for their salience, as indicated by the magnitude of their positive or negative discrepancy from the mean. Although the trend was in accord with the prediction, the hypothesis was not confirmed. For several girls there appeared to be a movement from interest in interaction and introception in elation to preoccupation with work in depression.

Hypothesis 6. Direction of Punitive Responses. The results of the Rosenzweig P-F test did not support the hypothesis concerning the direction of punitive responses. Contrary to our prediction, the frequency of extrapunitive responses was significantly higher for this group in depression (11 of 14 *S*s, $p < .05$). Intropunitive responses were significantly less frequent for the group in depression (11 of 14 *S*s, $p < .05$). Impunitive responses tended to be less frequent in depression than in elation (9 of 13 *S*s, 1 tie, not significant).

These results suggest that most *S*s, in satisfied and elated moods, are less hostile towards external objects, but that in depression they have a tendency to project more extrapunitive responses.

Hypothesis 7. Persistence of Effort. As predicted, the frequencies of obstacle-dominated and ego-defensive responses to frustration elicited by the Rosenzweig P-F test were higher in depression, but not significantly so. The prediction that the frequency of need-persistent responses would be lower in depression than elation was significantly supported (11 of 14 *S*s, $p < .05$). This set of findings is congruent with formulations such as Bull's which analyze depression as perseveration in helplessness and protest by a person who has become fixated on his own entanglements and discomforts in the face of frustrating interruptions of progress toward goals (Bull, 1951; Bull & Strongin, 1956).

Characteristics and Parameters of Mood Fluctuation

Mean Value of Average Mood. The individual means of the daily ratings of "average mood for the day" over the entire period varied from 5.43 to 7.37 for the 21 *S*s. The mean of their means was 6.14, and the median was also 6.14. The midpoint of the mood scale is 5.5, half way between "5. Feeling a little bit low, just so-so," and "6. Feeling pretty good, OK." With but one exception, the mean of the average mood of each *S* fell above 5.5;

in other words they tended more toward "elated" than "depressed" moods. The data are consistent with a substantial body of findings regarding avowals of "average level of happiness" (Goldings, 1954; Wessman, 1956).

Intensity of Mood Fluctuation. Two possible measures were considered for ascertaining intensity of mood swings. One was the variability of the "daily average" mood, as measured by its standard deviation, which indicated the degree of fluctuation of the average daily mood from its mean. This measure ranged from .42 to 1.81 in the group, with a median of .94. The other measure considered was the average range of feeling between the "lowest mood" and "highest mood" for each *S* for each day; thus it represents a measure based on the extremes of feeling of the *S* for each day. This "average daily range" varied from 1.17 to 4.36 for the group, with a median of 2.61.

To ascertain whether these two measures were comparable, whether those *S*s whose average mood fluctuated a great deal from day to day were also the *S*s given to the widest ranges of mood alteration within single days, the two were rank ordered and correlated. The result, though not statistically significant, may indicate some degree of positive correlation ($r = .34$). As both measures of intensity are of interest, both were used and are reported in the data analysis.

Our data did not confirm Flügel's (1925) inference that those who "experience the most extreme degrees of feeling are on the whole less happy than those whose feelings are usually less intense." The rank order correlation of the mean value of the average mood with the standard deviation was −.17, and with average daily range was +.36, neither of which is significant.

To see if there was a relation between the tendency to have high extremes and the tendency to have low extremes *S*s were ranked on (*a*) the average of the 5 "lowest" scores on the "lowest moods" and (*b*) the average of the 5 "highest" scores on the "highest moods." The rank orders correlated .49 ($p < .025$), indicating the *S*s who reported the most depressed moods also tended to be those who reported the most elated moods.

Frequency of Pronounced Mood Fluctuations. Alternative ways of measuring frequency of fluctuation were considered. A count of the number of times an *S*'s "daily average mood" crossed its mean was considered and rejected because it would make no distinction between relatively minor fluctuations about the mean and more pronounced alterations of mood—and it is usually the latter which is referred to when characterizing an individual as "moody" or "changeable." The measure of frequency adopted instead was a count of the number of periods that an *S*'s "average daily mood" was outside of the values 5, 6, and 7 on the mood scale (the values within which most of the fluctuations occurred). Thereby every *period* (one day or several days) the *S*'s mood was "8. Elated and in high spirits" or more extreme on the elated side, or "4. Spirits low and somewhat 'blue'" or more extreme on the depressed side, was counted as one. Thus this was a measure of the *frequency of pronounced mood swings.* Comparing these scores with visual inspection of the *S*'s records indicated it to be a satisfactory and representative measure of the phenomena. The *S*s ranged from 1 to 12 pronounced peaks and/or troughs over the 42-day period.

Patterning of Mood Fluctuation. Inspecting the plots of the *S*s' daily reports, it is apparent, as it was in Johnson's study (1937), that the fluctuations have more of an m shape than a w shape. That is, the periods of depression are generally represented by a sharp dip in the graph, usually of short duration, often only a day; the periods of relative elation continue over longer periods of time. This finding corresponds with the finding that the *S*s report feeling, on the whole, more often relatively elated than depressed.

Contrary to Johnson's interpretation of her observations, we did not observe any particular rhythm to our *S*s' fluctuations. Rather the periodicity and duration of fluctuations was decidedly irregular for all but one of the *S*s.

Relation of Menstruation and Mood. Johnson's study (1937) reported finding no apparent relationships of mood fluctuation with menstrual periodicity. This contradicts both folklore and investigations like Benedek and Rubenstein's (1942) which found feelings of frustration and irritability more prominent a few days before the menstrual period. Our data showed 14 out of the 19 *S*s reporting menstrual periods had an "average" mood lower than their general mean during the 2 days prior to menstruation ($p < .05$), and 8 of the 19 reached their lowest moods of the entire 6-week study on these 2 days. Of the 19 *S*s, 11 also showed an "average" mood lower than their mean on the day the menstrual period commenced.

Concomitants of Interindividual Differences on the Parameters of Mood Fluctuation

Hypothesis 8. Intensity and Frequency of Hedonic Fluctuation and Relative Consistency of the Self Concept. The correlations between the self descriptions provided in depression and in elation, which ranged from .18 to .79, with a median of .63, provided the basis for a rank order of the 14 *S*s on relative consistency of the self concept. Rank order correlations were then computed against ranks on intensity of fluctuation and on frequency of pronounced mood fluctuations. Both parts of the hypothesis were significantly confirmed. Self consistency was negatively related to both measures of intensity of fluctuation, the standard deviation of the daily average ($r = -.64$, $p < .01$) and the average daily range ($r = -.46$, $p < .05$). Self consistency was also negatively related to frequency of pronounced mood fluctuations ($r = -.63$, $p < .01$). These data support the hypothesis that girls who have relatively intense mood changes, or relatively frequent ones, have less stable conceptions of their own identity. Average mood level, on the other hand, was unrelated to self consistency ($r = .09$).

It was shown above (Hypothesis 2) that the ideal concept changed less than the self concept with changes in mood. The small changes which did occur in the ideals were not significantly related to the intensity, frequency, or average level of the *S*s' mood fluctuations.

Hypothesis 9. Average Hedonic Level and Relative Self Esteem. The measure of self esteem was the one used in Hypothesis 1, the correlation between the self and ideal descriptions. The *S*s were rank ordered on these correlations in both elation and depression, and these ranks were correlated against a rank order on the mean "average daily mood" over the six-week testing period. Contrary to prediction, the correlations were negative (in elation $r = -.36$, not significant; in depression $r = -.50$, $p < .05$).

No significant relationships between either intensity or frequency of mood fluctuation and relative degree of self esteem were found for either elated or depressed states. Rank on self esteem in depression was only moderately correlated with self esteem in elation ($r = .40$, not significant).

The high negative correlations found between self esteem (correlation between self and ideal Q sorts) and happiness (mean "average daily mood") required further exploration. Two explanations seemed possible: 1. The happiness score might be in error, or reflect some quality other than happiness. Error seemed unlikely, since the scale was based on a six-week average of the "average daily mood" score and was therefore unlikely to have been greatly influenced by transient moods, misunderstanding of the scale, etc., and because an analysis of variance had shown that the between-subject variance, as compared to that within the individual daily records, was significant well beyond the .01 level. It is possible, however, that the average position on the happiness scale might be influenced by characteristics other than happiness, e.g., response sets to score high or low, or a set in some *S*s to deny unhappiness. 2. The self-ideal correlation might measure qualities other than self esteem. A high correlation between the self and the ideal might reflect denial of undesirable qualities in the self (Block & Thomas, 1955), a form of ego restriction which might be more typical of unhappy people than of happy ones. Or the same high correlation might reflect restriction of aspirations, disgruntled acceptance of limitations, and lowering of ideals.

The content of the Q-sort self and ideal descriptions was examined for evidence bearing on the possibilities outlined above. The group of *S*s was divided in half according to their mean "daily average mood" over the six-week observation period, and the Q-sort descriptions provided by the seven relatively more happy girls were compared with the descriptions provided by the seven relatively less happy girls. The typical self and ideal characterizations of the "happy" and the "unhappy" girls are given below. Then the significant discrepancies between the self and ideal descriptions are summarized and discussed, first for the happy and unhappy girls in elation, then for both groups in their relatively depressed moods.

Self Descriptions. The composite self picture presented by the happy girls, both in depression and in elation, was that of a friendly, altruistic person with diverse interests and some sense of perspective about herself, able to enjoy work without being preoccupied with it, and not considering herself sophisticated or poised. The change from elated to depressed moods seemed to be characterized by a decrease in breadth of interests and in satisfaction with learning, and an increase in nervousness, worried concern over school work, introspection, and concern with emotional security.

The unhappy girls described themselves, in both moods, as introspective, altruistic, idealistic, sympathetic and understanding of the problems of others, and concerned with academic work.

Comparison of the self descriptions of the two groups indicated that the unhappy girls described themselves as more introspective, independent, sophisticated, self-sufficient, critical, and more interested in academic work; the happier girls described themselves as more friendly, sociable, unorganized, nervous, tolerant, more willing to give of themselves to others, and more interested in what others had to offer.

Ideal Descriptions. In their elated moods both groups had ideals which were alike in most respects, emphasizing qualities of being interesting and stimulating, having a good sense of humor about themselves, and being sensitive and perceptive of the feelings of others. The happy girls tended to place a higher value on friendliness and consideration for others, while the unhappy girls placed a higher value on work.

In depression both groups valued work more than in elation, but this emphasis on work was much more extreme for the unhappy girls. For the happy girls concern with work in depression took the form of putting a higher value on the inherent satisfactions in learning and greater concern with discovering what they wanted out of life; for the unhappy girls it took the form of more exclusive concern with intellectual creativity and achievements.

The data indicate that one of the suggested explanations for the high self-ideal correlations of the unhappy girls, that they had lowered their ideals, is not valid. It appears more probable that the unhappy girls had *narrowed* their ideals, particularly those ideals they described in moods of depression, to exclude social values and place a more exclusive emphasis on scholastic and intellectual achievement.

Discrepancies between Self and Ideal. Examination of those items for which there were significantly consistent discrepancies (sign test) between selves and ideals showed 10 such items for the happy and 9 for the unhappy girls in depressed moods, 5 for the happy and 0 for the unhappy in elated moods. These differences indicate that in depression both groups experienced discrepancies between what they were and what they would like to be but that the unhappy girls, as a group, felt no consistent divergence from their ideals in elation, while the happy girls continued to feel areas of discontent. For the happy girls most of the discrepancies between actuality and aspiration concerned social and emotional qualities: ability to be interesting and stimulating, poise and self-confidence, and emotional security. The unhappy girls, in contrast, experienced discrepancies primarily in terms of unfulfilled intellectual ambitions. Elation seemed to be experienced by the unhappy girls as satisfaction of these internal standards, while for the happy girls elation was accompanied by outgoing interest in others and desire for the qualities which would enable this interest to be realized.

DISCUSSION

Intra-individual variation in psychological measurements has generally been ignored or regarded as a statistical inconvenience; a comparatively small number of investigations has made such variations a subject of study (Fiske & Rice, 1955) and only a few psychologists have stressed the value of methods of repeated measurement, or systematically employed them (Horn, 1950; Cattell, 1957). It is clear that repeated measurements play a basic role in other fields, particularly physiology and economics, and it is equally clear that some psychological characteristics, such as moods, vary markedly within individuals as well as between them. One aim of this study was the development of techniques for repeated measurements of moods and related feelings. The "Mood scale" used in this investigation has not yet been shown to have such desirable scale attributes as unidimensionality, equality of intervals, interindividual comparability, etc. For the present we can report that the scale seemed clear and relevant to the *S*s, and that their changes on the scale were related to demonstrable differences in self conception and projective reactions to frustration. Further study of this scale, and of the other scales whose details are not considered here, is currently in progress.

It should be recognized that the moods and self conceptions reported here are conscious avowals of subjective feelings and beliefs, and their validity is limited to an unknown extent by possible deficiencies in some *S*s' self awareness and capacity to experience, and perhaps by unwillingness or inability to report subjective states accurately. The data show only some aspects of the phenomenology of moods, and only by inference can this material be related to the unconscious dynamics central to affective life.

Although 25 *S*s commenced the study, the tests of the hypotheses are based only on the 14 *S*s who provided sufficiently complete data; there may be some systematic bias introduced by the omission of those who did not provide complete records. The use of *S*s from a homogeneous but relatively unstudied environment probably reduced extraneous and irrelevant variability within the sample to a near minimum; but it also made uncertain the degree to which some of the relationships found might reflect particular factors in the Harvard-Radcliffe subculture. The differences we found between the relatively happy and relatively unhappy girls may not indicate consistent differences between happy and unhappy people generally but, rather, reflect the particular role conflicts and strains at Radcliffe or similar women's colleges. It appeared that in this group the less happy girls had identified more with a "student" or "scholar" role, the more happy girls with a "date" or "prospective wife" role.

The major concern of the study was to test the general hypothesis of continuity and similarity of dynamic patterns between normal and pathological moods. Pathological moods have been more thoroughly studied and reported than normal moods, and there is substantial agreement as to their characteristics. A fairly broad and representative set of generalizations could therefore be drawn from the literature dealing with affective psychopathology and stated as hypotheses. Most of the findings which compared elated and depressed states in the same girls (i.e., intra-individual changes) supported the continuity hypothesis. Self esteem, as measured by the self-ideal correlation, was shown to be lower in depression than in elation, and the change was shown to be due mainly to a drop in feelings of goodness and worth in the self, with a possible secondary cause being an increase in the stereotyped "goodness," and presumably demanding character, of the ideal. Responses to the Rosenzweig P-F test indicated, as predicted, that the girls were less likely to go ahead to overcome obstacles in depression than in elation. One deduction from the continuity hypothesis was significantly reversed: *S*s were not, as had been predicted, more intropunitive in depression (Rosenzweig P-F test); rather they tended to be more punitive generally and were significantly more extrapunitive. There is general consensus in the clinical literature that anger and aggression are predominantly turned inward in severe depression. Our results suggest that the point at which mounting hostility shifts from an outward to an inward direction may mark the transition from relatively mild depression to severe affective pathology.

Comparison of the *S*s with each other (interindividual comparison) on major parameters of mood change indicated that girls with stable self conceptions were less prone to experience either pronounced or frequent mood fluctuations. Priority of cause and effect relationships cannot be established here; probably individuals who have experienced marked and frequent changes in mood, either from physiological causes or due to developmental traumas, do not develop stable identities, and in later life the lack of a stable identity makes then more liable to such emotional storms.

A third interindividual comparison found that individuals with high self esteem (as measured by self-ideal correlation) were relatively unhappy, while individuals with low self esteem were relatively happy. Initially it may seem that this finding contradicts the earlier finding that *S*s had higher levels of self esteem when they were elated and lower levels of self esteem when they were depressed. The earlier finding, however, was based on an intra-individual comparison (each girl in elation compared with herself in depression), while the latter finding is based on an interindividual comparison of girls who were characteristically more happy with other girls who were characteristically somewhat less happy. It is logically fallacious to assume that relationships which can be demonstrated in intra-individual comparisons necessarily hold in interindividual comparisons, and vice versa; but often the experimental convenience of interindividual comparisons (all *S*s tested at one sitting, etc.) causes psychologists to overlook this limitation, and to argue that group comparisons support or contradict hypotheses which could be directly tested only by intra-individual techniques.

Even taking into account the logical differences between inter- and intra-individual tests of hypotheses, the finding of a negative correlation between self esteem and hedonic level still seemed paradoxical and required exploration of the content of the self and ideal descriptions as related to characteristic hedonic level. When the self and ideal descriptions of the more happy and less happy girls were compared they were found to show two contrasting but internally consistent patterns of feelings, interests, and aspirations. The unhappy girls resembled Hamilton and Mann's description of the predepressive personalities of patients with involutional psychosis (melancholia) who were found "to take life very seriously, to follow rigid patterns of behavior, to display lack of imagination and a sense of humor," and to be unsociable, lack spontaneity, have patterns of careful intellectual control, and to command respect "because of their thoroughness, conscientiousness, meticulous devotion to duty, and above-average level of productiveness." (Hoch & Zubin, 1954, p. 201). While it would be ridiculous to regard half of the *S*s as predepressive, this close correspondence does suggest that there is considerable resemblance between the personalities of the girls whose average level of happiness is below the median in this group and the personalities of those who eventually manifest depressions of pathological depth.

SUMMARY

A set of 10-point self-rating affective scales was developed for repeated daily measurements over extended periods. Records were kept over a 42-day period by 14 female college students who also took two identical sets of psychological tests (self and ideal Q-sort descriptions, Rosenzweig P-F test), once when they were elated, once when they were depressed. Hypotheses were tested as to intra-individual changes in elation and depression and as to concomitants of interindividual differences on the parameters of mood fluctuation. The significant intra-individual changes in depression (relative to elation) were: lower correlation of self and ideal, more unfavorable self concept, increased *extra*punitive responses; and reduced need-persistive responses. Individual records of mood fluctuation were markedly irregular and showed no clear periodicity. Generally the average mood was slightly on the elated side, and depressed periods were of short duration. The days prior to onset of menstruation were generally depressed. Relative interindividual standings on the parameters of mood fluctuations showed

the following significant correlations: intensity of mood swings inversely related to relative stability of the self concept; frequency of pronounced mood swings inversely related to relative stability of the self concept; and average mood (or general hedonic level) inversely related to correspondence of self and ideal. The last finding, which was contrary to prediction, led to an examination of the contents and discrepancies between self and ideal descriptions in elation and depression for the relatively more and relatively less happy girls. The less happy girls appeared to put more stress on intellectual achievements and their periods of elation seemed to correspond with meeting their high standards, while the happier girls appeared to have more diverse interests and be more sociable and even in elation experienced discrepancy between self and ideal. The relevance of these results to a general hypothesis of continuity between normal and pathological functioning was considered. The results suggest the value for personality research of systematic measurement of emotional fluctuations over extended periods, and of the consideration of affective states in personality testing.

REFERENCES

ABRAHAM, K. Notes on the psychoanalytical investigation and treatment of manic-depressive insanity and allied conditions (1911). In *Selected papers of Karl Abraham.* New York: Basic Books, 1954. (a)

ABRAHAM, K. A short study of the development of the libido, viewed in the light of mental disorders (1924). In *Selected papers of Karl Abraham.* New York: Basic Books, 1954. (b)

BELLAK, L. *Manic-depressive psychosis and allied disorders.* New York: Grune & Stratton, 1952.

BENEDEK, T., & RUBENSTEIN, B. B. The sexual cycle in women, the relation between ovarian function and psychodynamic processes. *Psychosom. Med. Monogr.*, 1942, **3,** Nos. 1 and 2.

BLOCK, J., & THOMAS, H. Is satisfaction with the self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51,** 254–259.

BULL, NINA. The attitude theory of emotion. *Nerv. ment. Dis. Monogr.*, 1951, No. 81.

BULL, NINA, & STRONGIN, E. The complex of frustration: A new interpretation. *J. nerv. ment. Dis.*, 1956, **123,** 531–535.

BURTON, R. *The anatomy of melancholy.* 1621.

CATTELL, R. B. *Personality and motivation structure and measurement.* Yonkers-on-Hudson: World Book, 1957.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FISKE, D. W., & RICE, LAURA. Intra-individual response variability. *Psychol. Bull.*, 1955, **52,** 217–250.

FLÜGEL, J. C. A quantitative study of feeling and emotion in everyday life. *Brit. J. Psychol.*, 1925, **15,** 318–355.

FREUD, S. Mourning and melancholia. (Originally published 1917). In *Collected papers.* Vol. 4. London: Hogarth, 1953.

GOLDINGS, H. J. On the avowal and projection of happiness. *J. Pers.*, 1954, **23,** 30–37.

HOCH, P. H., & ZUBIN, J. (Eds.) *Depression.* New York: Grune & Stratton, 1954.

HORN, D. Intra-individual variation in the study of personality. *J. clin. Psychol.*, 1950, **6,** 43–47.

JOHNSON, WINIFRED B. Euphoric and depressed moods in normal subjects. *Character & Pers.*, 1937, **6,** 79–98, 188–202.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 1. *Theory and method.* Cambridge, Mass. Addison-Wesley, 1954. Pp. 289–334.

RADO, S. The problem of melancholia. *Int. J. Psychoanal.*, 1928, **9,** 420–438.

ROSENZWEIG, S., FLEMING, EDITH E., & CLARKE, HELEN J. Revised scoring manual for the Rosenzweig Picture-Frustration study. *J. Psychol.*, 1947, **24,** 165–208.

WESSMAN, A. E. A psychological inquiry into satisfactions and happiness. Unpublished doctoral dissertation, Princeton Univer., 1956. (Mic. 57-2228, University Microfilms, Ann Arbor, Mich.)

(Received February 5, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

REINFORCEMENT OF VERBAL AFFECT RESPONSES OF NORMAL SUBJECTS DURING THE INTERVIEW[1]

KURT SALZINGER AND STEPHANIE PISONI

New York State Department of Mental Hygiene and Biometrics Research, Columbia University

The recent application of behavior theory to the study of verbal behavior has stimulated a great deal of research as is well evidenced in two recent reviews of the area (Krasner, 1958; Salzinger, K., 1959). The extension of this approach to the experimental analysis of the interview has been conceptualized by the senior author in a recent paper (K. Salzinger, mimeo edition, 1959) and has met with some success in actual experiment. So far, it has been shown that the output of verbal affect responses of schizophrenics can be manipulated by the use of discriminative stimuli such as questions (Suzanne Salzinger, 1956) and by the use of reinforcements like verbal expressions of agreement (Suzanne Salzinger, 1956; K. Salzinger & Pisoni, 1957, 1958a, 1958b). In addition, it has been shown that the effectiveness of conditioning is independent of the placement of the conditioning period in the interview, i.e., at the beginning or in the middle (K. Salzinger & Pisoni, 1957). Finally, the rate of extinction of affect responses has been found to vary directly as a function of the rate of acquisition (K. Salzinger & Pisoni, 1958b).

The present study attempts to extend the above findings by conditioning affect statements in a sample of normal individuals and comparing the results obtained with those of a sample of schizophrenics.

METHOD

Subjects

Eighteen female and eight male normal *S*s ranging in age from 20 to 50 years with a median age of 37.5 were selected from among the admissions to a general medical hospital. Patients with obvious psychosomatic symptoms such as ulcers and hypertension and those who showed evidence of any history or gross current signs of psychosis were excluded from the sample. Included were patients with illnesses such as pneumonia, tumors, nephrosis, gall bladder disorders, yellow jaundice, arthritis, and appendicitis, and those requiring minor surgical operations.

Experimental Procedure

Each *S* was interviewed for a period of 30 minutes. Thirteen *S*s were interviewed by a female *E* and 13 by a male *E*. All interviews were recorded with the recording equipment placed in full view of the patient. *S*s were informed that they had been selected randomly and that these interviews were being conducted in order to determine how such an interview might be incorporated into the routine medical examination administered in the hospital.

The interview began with specific questions concerning the name of the patient, his age, marital status, number of children and number of siblings. All the patient's responses to these questions were reinforced by *E*'s repetition of the response given, by his writing it down, and by his emission of phrases such as "mmm-hum," "uhhuh," "I see," etc. This procedure was adopted in order to obtain factual information upon which subsequent interview questions could be based and to establish the interviewer as a source of reinforcement, in this way encouraging the patient to speak in the presence of *E*. Following this, the experimental part of the interview was initiated with the question: "Would you tell me why you are here in this hospital?" Subsequently, the interviewer asked additional general questions whenever the patient ceased talking for more than two seconds.

Each interview was divided into three 10-min. periods. During the first 10 min. (operant level), the base rate of affect responses was determined. *E* asked questions but did not reinforce any statement made by the patient. Reinforcement was defined as verbal agreement consisting of such phrases as "mmm-hum," "uhhuh," "I see," and "yeah," immediately following statements made by the patient. During the second 10 min. (conditioning), *E* continued to question the patient and, in addition, reinforced each affect response by immediately following it with verbal agreement. During the last 10 min. (extinction), *E* withheld all reinforcement but continued asking questions whenever they were necessary to maintain the flow of speech of the patient.

Definition of Response

The response class of affect used in this experiment was defined as any statement by the patient describing or evaluating his own state (other than intellectual or physiological). The response class therefore included only those affect statements which begin with the pronouns "I" or "we" as the following examples indicate: I am satisfied, I'm happy, we enjoyed it, I like him, I'm very close to him, I was mad at him, we hated him, I'll always be jealous of him, I am upset, I am a lonely person, I was so ashamed, I'm sorry for him, I feel (followed by any words), I was frightened, we couldn't take it, I always suffered. Quotations in which affect was ascribed to the speaker by another individual (e.g., "My husband said I didn't feel good"), although fitting all other criteria, were excluded on the basis of not being direct expressions of the *patient's* affect.

---

[1] This study was supported by Research Grants M586 (C3) and M1541 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The authors are indebted to J. Zubin and S. Sutton for their constant help and interest in this study. The authors are also grateful to L. C. Kolb and R. Senescu, who made the normal population available for this study, and to B. Schoenberg, who very generously gave of his time in the selection of *S*s. This paper was presented in part at a meeting of the Eastern Psychological Association, Philadelphia, Pa., 1958.

Statements like "I am happy and excited" were considered as one affect statement only, because the pronouns "I" or "we" did not precede the second affective word. On the other hand, incomplete statements (in the sense that the object of the affect is not mentioned) like "I love… " or "we feared… " were viewed as bona fide responses.

Certain types of private events or internal states were excluded from the response class of affect because they referred primarily to intellectual processes, to actions which are sometimes but not always associated with affect, or to desires which appear to constitute a class of responses different from the affect class as defined here. Some examples are the following: I am confused, I am confident, I would like to…, I want, I was surprised, I am not well, we forgave him, I threaten her constantly, I didn't trust them.

The reliability of coding by the interviewers was examined in an earlier study (K. Salzinger & Pisoni, 1958a), yielding proportions of agreement ranging from .79 to 1.00, showing that both *E*s would generally have reinforced the same responses in identical interviews. Comparable coefficients of agreement were obtained for the interviews of this study.

A count per minute was taken of statements belonging to the general class of self-references (statements beginning with "I" or "we" which included among them self-referred affect statements) in order to compare changes in the occurrence of this class with those of the class of self-referred affect statements.

Interviewer Questions

Some or all of the following topics were discussed during each interview: reasons for being in the hospital and causes for illness; patient's relationships to his parents, siblings, fellow employees, employers, fellow students and teachers, wife or husband, children, friends; patient's activities during free time, and plans for the future. *E* made an attempt to balance these topics over the three conditions. For instance, if the patient discussed the symptoms of his illness in the operant level condition, *E* asked questions regarding the possible causes of the illness during conditioning and brought the patient back to these topics during extinction. As long as the topics were approximately balanced over the three conditions, however, *E* took his cues for topics from the content of the patient's statements. Both the number of topics and the order in which they were discussed varied from interview to interview.

Questions designed for eliciting affect directly, such as, "How did you feel about that?" or "Were you happy?" were not used.

RESULTS

Conditioning Effect for the Entire Group

The difference in number of affect statements emitted during operant level, conditioning, and extinction was tested by Friedman's nonparametric analysis of variance (Siegel, 1956) and found to be statistically significant ($p = .05$). The greatest number of affect statements was found during conditioning (sum of ranks = 60.5), the next greatest during extinction (sum of ranks = 52.5), and the smallest during operant level (sum of ranks = 43.0).

When the same statistical procedure was applied to general self-referred statements, no statistically significant difference was found ($p = .72$) among the three periods. Examination of the trends revealed that the largest number of self-referred statements occurred during operant level (sum of ranks = 55.0), the next largest during extinction (sum of ranks = 51.5), and the smallest during conditioning (sum of ranks = 49.5).

Conditioning Effect for Subgroups

In previous studies (K. Salzinger & Pisoni, 1957; 1958b), it was demonstrated that the conditioning effect takes place if *S* is given more than a certain minimum number of reinforcements. *S*s who received less than the minimum evidenced no changes in behavior as a result of the delivery of reinforcement. This finding held when a given number of reinforcements was assigned to each *S* before the interview as well as when reinforcements were made available only for a given length of time. When the present group of normals likewise was divided into those *S*s who received less than 8 reinforcements and those who received 8 or more, a significant difference in the number of affect statements between the conditions of operant level, conditioning, and extinction was found for the high reinforcement group ($p < .01$) but not for the low reinforcement group ($p = .90$). The high reinforcement group revealed the greatest number of affect statements during conditioning (sum of ranks = 32.5), the next greatest during extinction (sum of ranks = 28.0), and the smallest during operant level (sum of ranks = 17.5). Examination of the general class of self-referred statements for the high and low reinforcement groups indicated no statistically significant difference among the three periods of the interview; the p level for the high reinforcement group was .90, for the low reinforcement group, .64.

Comparison of Normal and Schizophrenic Subjects

The normal *S*s of this sample were compared to a sample of schizophrenic patients from a previous study (K. Salzinger & Pisoni, 1958a). These schizophrenics had been subjected to the same procedure as were the present *S*s. Comparisons were made separately on each condition by the Mann-Whitney U test (Siegel, 1958). No statistically significant difference was found. The p value for the difference between operant levels was .20, for conditioning, .17, and for extinction, .87.

Members of the normal group were then matched with members of the schizophrenic group in terms of the following variables: sex, operant level, and number of reinforcements received. A pair was considered matched when the members were of the same sex, when the number of reinforcements received by each member differed by no more than 2, and when the discrepancy between the operant levels was no greater than three affect statements. Twelve pairs of individuals were found to meet these requirements. Although *S*s were not matched on rate of acquisition, application of the paired replicates test (Siegel, 1956) to the normal–schizophrenic pairs revealed no significant difference ($p = .10$) with respect to this variable. The matched normals and schizophrenics were then compared on their resistance to extinction.

Application of the paired replicates test to these pairs revealed greater resistance to extinction in normals than in schizophrenics ($.02 < p < .05$). Nine of the normals extinguished more slowly than the schizophrenics; two showed no difference, and one schizophrenic extinguished more slowly than his normal match. The median difference in number of extinction responses was 6, the normals emitting 6 more affect responses, on the average, than did the schizophrenics during extinction.

The same comparison of individuals matched on operant level alone or on number of reinforcements alone did not serve to differentiate normals from schizophrenics ($p > .10$). Figure 1 represents individual cumulative curves of affect responses for two pairs of matched schizophrenics and normals over the three conditions. One matched pair has a low operant level, the other a high operant level. Inspection of this figure reveals both the similarity of the operant level and conditioning rates for the matched pairs as well as the difference in their extinction rates.

DISCUSSION

The results of this study resemble those previously obtained (K. Salzinger & Pisoni, 1958a). The interviewer's verbal reinforcement increased the frequency of uttering affect statements. Data taken from subgroups make it clear that a certain minimum number of reinforcements is necessary for conditioning to take place; i.e., an *S* who is given less than this number generally shows no evidence of conditioning. Investigation of the frequency of general self-referred statements showed it to be invariant over the three conditions of the interview, thus yielding evidence for the specificity of the effect of reinforcement upon self-referred affect statements.

While the finding that normals show greater

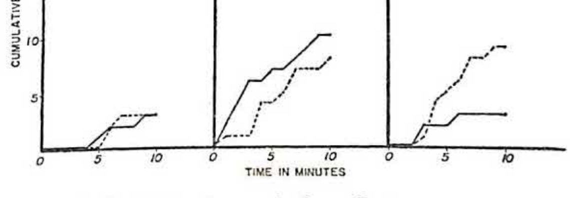


FIG. 1. Individual cumulative affect response curves for two pairs of matched schizophrenics and normals. The top pair shows two subjects with high operant levels, while the bottom pair shows two subjects with low operant levels.

resistance to extinction than schizophrenics is in agreement with the clinical observation of greater affect in normals, it suggests that the clinically "recognized" differences in amount of affect expressed are primarily a function of *S*s' differential reaction to the cessation of reinforcement. It now becomes necessary to see whether this difference in extinction holds only for the response class of affect or whether it applies to any response class. Such a study is presently in progress.

The difference in extinction rate has an important implication for psychotherapy with schizophrenics if the application of reinforcement is a useful tool in therapy. Some evidence for this contention has been made available by Murray (1956). The implication is that even though schizophrenics may appear to profit from psychotherapy (administration of reinforcement may result in conditioning), the profit (conditioning effect) may not persist as long in such patients as it does in normals after the therapy (reinforcement) is terminated. It remains for further studies to determine how much larger the number of verbal reinforcements must be for schizophrenics than for normals in order for the former to gain as much therapeutic profit from reinforcement as the latter.

SUMMARY

Twenty-six normal individuals, hospitalized for a variety of physical complaints, were interviewed for a period of 30 minutes. Each interview consisted of a 10-minute operant level period, during which *E* asked questions necessary to maintain *S*'s talk but did not otherwise respond to *S*'s speech; 10 minutes of conditioning, during which *E* reinforced by agreement all self-referred affect statements; and 10 minutes of extinction, during which *E* withheld all reinforcement. Conditioning of the verbal response class of self-referred affect statements occurred, and normal *S*s showed greater resistance to extinction than did schizophrenics.

REFERENCES

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

MURRAY, E. J. A content-analysis method for studying psychotherapy. *Psychol. Monogr.*, 1956, **70** (13, Whole No. 420).

SALZINGER, K. The experimental approach to the interview. In J. Zubin (Ed.), *Experimental abnormal psychology.* Available in mimeo form at Columbia Univer. Library, 1959.

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61**, 65–94.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement verbal affect responses of schizophrenics during the clinical interview: The effect on conditioning of placement of the period of reinforcement. Paper read at the American Psychological Association, New York City, 1957.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, **57**, 84–90. (a)

SALZINGER, K., & PISONI, STEPHANIE. Verbal affect responses of schizophrenics as a function of number of reinforcements and rate of acquisition. Paper read at the Eastern Psychological Association, Philadelphia, Pa., 1958. (b)

SALZINGER, SUZANNE. Rate of affect response in schizophrenics as a function of three types of interviewer verbal behavior. Paper read at Eastern Psychological Ass., Atlantic City, March, 1956.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received February 13, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

CHANGE IN ATTITUDES TOWARD THE NEGRO IN A SOUTHERN UNIVERSITY

ROBERT K. YOUNG, WILLIAM M. BENSON, AND WAYNE H. HOLTZMAN
University of Texas

A FEW months after the momentous 1954 United States Supreme Court decision concerning segregation an attitude survey was conducted among undergraduate students at the University of Texas (Kelly, Ferson, & Holtzman, 1958). Included in this questionnaire was the 26-item Desegregation Scale (D scale) for measuring attitudes toward the Negro. Background factors found by Kelly et al. (1958) to be significantly related to attitude toward the Negro included the person's geographic region of origin, father's income, major field of study, religious preference, and stated frequency of church attendance.

The existence of these baseline data presented a unique opportunity to study the degree of change, if any, in attitudes toward the Negro after a lapse of several years. The original study had employed a large, stratified random sample of 547 cases which could be compared in detail with an identical sample of undergraduates given the same questionnaire in the winter of 1958. During the intervening three-year period much had happened on the American scene with respect to school desegregation. Stiffening resistance through-out the South came to a dramatic climax in the integration crisis at Little Rock, Arkansas, in late 1957 when Federal troops were called out by President Eisenhower to force the opening of Central High School. The voice of the Southern moderate was drowned out completely by extremists on both sides of the issue.

The period immediately following the Little Rock crisis represented an ideal time for repeating the earlier survey to determine the nature of change in attitudes toward the Negro. The undergraduate student body at the University of Texas is an excellent group for such a study, since cultural mores and life styles characteristic of the deep South, the West, the North, and Mexico are all represented.

PROCEDURE

Great care was taken to duplicate the procedures used in the 1955 survey so that the two sets of data would be comparable. Date cards from the Registrar's Office were used to draw a 5%, random stratified sample of native born, full-time undergraduate students. Stratification was based on sex and religious preference. Of the 629 students to whom letters were mailed, 265 came in immediately to fill out the attitude questionnaire in previously announced group sessions. Another 151 completed the questionnaire after a telephone follow-up. Questionnaires were mailed to the remaining subjects, and 101 of these responded, making a total of 517, or 82% of the original sample (comparing favorably with the 84% obtained in 1955). Preliminary analysis of the data revealed no bias in the questionnaires returned by mail. Since 55 of the 112 nonrespondents had dropped out of school prior to the study, the actual percentage of returns was 91. Eliminating 20 questionnaires that were improperly answered left a total of 497 cases for analysis.

The 90-item attitude questionnaire was given anonymously and scored in the same manner as reported by Kelly et al. (1958). The D scale had been formed from data in a preliminary scaling study by taking the 26 items with minimum variance and approximately equal interval distribution on the 11-point scale. No items were included with scale values ranging from 4.4 to 7.2. Each of the items was scored according to the Likert weighting procedures with extremely positive responses scored zero and extremely negative ones scored four. Summing across all 26 items gave a possible range of scores from zero to 104.

RESULTS

Over-all Comparisons of the 1955 and 1958 Samples

Of paramount interest are differences in the two samples which would indicate any significant shifts in attitude toward the Negro during the three-year interval between 1955 and 1958. The distributions of scores on the D scale were markedly similar in the two samples. The means were 46.2 and 45.3 for the 1955 and 1958 samples, respectively. Similar values for the standard deviations of the distributions were 18.8 and 18.7. Clearly there are no measurable changes in degree of acceptance of the Negro for the undergraduate student body as a whole.

In view of the widespread belief that the so-called moderates have been forced to move away from a neutral position because of the aggressive action of extremists, a comparison of the variances in the two samples is of special interest. The stiffening of extremes should lead to a significant increase in the variance of scores on the D scale. Actually the variance *decreased* slightly from 353 in 1955 to 348 in 1958. These results suggest that the social conflict characteristic of the period from 1955 to 1958 had little or no effect upon attitudes toward the Negro in Texas.

As a further test of changes that might have taken place during the three-year interval between the two surveys, an analysis of variance, corrected for disproportionate cell frequencies, was computed in which the main effects were (*a*) year of survey, (*b*) sex of respondent, and (*c*) fraternity (or sorority) membership. It is quite possible that significant interactions could take place between these factors even though the over-all distributions for the two surveys were markedly similar.

Two significant differences were observed. The first was fraternity affiliation. An F of 4.04 ($P < .05$) was obtained between respondents who belonged to a fraternity or sorority and those who did not. The unaffiliated students tended to be more tolerant of the Negro, a trend that was more pronounced in 1958 than in 1955 (although the sample-by-fraternity interaction failed to prove significant).

The second significant difference was the only one found involving the year of survey as a factor. A significant sex-by-sample interaction ($F = 4.34$, $P < .05$) indicated that the attitudes of men and women changed during the three-year period between surveys but in *opposite directions*. Inspection of Table 1 reveals that men were less tolerant of the Negro in 1958 than in 1955, while women showed a significant increase in tolerance.

TABLE I

CHANGE IN ATTITUDE TOWARD THE NEGRO FROM 1955 TO 1958, MEAN SCORES ON THE D SCALE

	Year of Survey	
	1955	1958
Men	44.8	47.9
Women	45.1	42.3

Additional Analysis of the 1958 Sample

As in the study by Kelly et al. (1958), a complete analysis was made of personal background factors and their relationship to attitude toward the Negro as revealed by the D scale. Since most of these results are in full agreement with the findings obtained for the 1955 sample, they are only briefly summarized. A more extended discussion has already been given in the earlier report. Mean D scale scores are cited parenthetically after the appropriate group.

Geographic location. The area from which the respondent comes is strikingly related to his attitude toward the Negro. In general, tolerance for the Negro increases as one moves from east Texas (mean score of 55.2) to west Texas (45.5) and the Rio Grande Valley (42.9).

Church affiliation. The Baptists (49.8) show the greatest amount of resistance to desegregation, with those stating no religious preference at the opposite extreme (32.4). The Catholics (44.3) and Jews (39.0) are the most tolerant of the major denominations analyzed.

Church attendance. Frequency of church attendance is related to attitude toward the Negro in a curvilinear fashion, again verifying an earlier theoretical deduction by Allport (1954). Those individuals who attend church every Sunday (44.3) and those who never attend church (38.1) are significantly more tolerant than those who claim they attend church once a month (51.4) or twice a month (48.4).

Grade point average. Grade point average is linearly related to D scale scores. Students who are academically superior—B average or above (42.0)—are more acceptant of the Negro than students with grades of C or below (48.6).

Major field of study. Social science majors (35.3) tend to be the most tolerant of the Negro, while students enrolled in pharmacy (57.2) or business (49.9) are the least tolerant.

Size of hometown. Students coming from cities with a population of 50,000 or over (43.3) are more tolerant of the Negro than those who reside in smaller towns and rural areas (48.1). While this same trend was noticed in 1955, it was not sufficiently pronounced at that time to be statistically significant.

Father's income. The higher the stated income, the less tolerant the individual. Students whose fathers had an income of less than $5,000 per year (41.9) were most tolerant, while those who claimed an income of over $20,000 (53.0) were least tolerant.

CONCLUSIONS

Attitude toward the Negro as measured by the D scale appears to be remarkably stable as far as group means are concerned. In view of the social tension prevalent in the South during the period 1955–1958, one might reasonably expect to find a change in attitude toward desegration, at least with respect to a widening of the breach between advocates of different stands on the issue. Contrary to such popular beliefs, no change in either mean score or variance of the D scale was found in comparing a random sample of students at the University of Texas in 1958 with a similar sample in 1955.

The only significant change in attitude over this three-year period is uncovered when men and women in the two samples are analyzed separately. While the women appear to be slightly more tolerant of the Negro in 1958, the men have moved in the opposite direction.

In both the 1955 and 1958 samples, an analysis was made of relationships between personal background characteristics of the student and attitudes toward the Negro. A striking degree of consistency was found across the two samples for such factors as geographic region from where the individual came, church affiliation, stated frequency of church attendance, and major field of study in college. Size of hometown and father's income were also significantly related to attitude score in 1958, although these same trends fell short of significant levels in the 1955 sample.

REFERENCES

ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass.: Addison-Wesley, 1954.

KELLY, J. G., FERSON, JEAN E., & HOLTZMAN, W. H. The measurement of attitudes toward the Negro in the South. *J. soc. Psychol.*, 1958, **48**, 305–317.

(Received August 17, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

CRITIQUE AND NOTES

SOME THEORY AND DATA CONCERNING THE ATTENUATION OF AGGRESSION[1]

KENNETH R. WURTZ
Michigan State University

When a child is punished for aggression he gradually learns to anticipate punishment whenever the aggressive drive is activated. This anticipation of punishment, occurring simultaneously with excitation of aggressive behavior, is anxiety. The child learns to be anxious just as he learns most of his other responses, and anxiety comes to be a drive state in and of itself. However, as a drive, anxiety is unique in one respect. When it occurs it is almost always in conflict with other drive states of the organism. In its essence, anxiety is a conflict-producing drive; when it is evoked by aggression, the anxiety-instigated responses conflict with those induced by the child's need to be aggressive.

From a systematic standpoint, anxiety is both an inhibitor and a stimulus. It is an inhibitor of overt forms of aggression by virtue of its capacity to instigate avoidance responses that conflict with aggression, and it is a stimulus to aggression since its occurrence commonly coincides with that of expressed aggression. Since the state of anxiety occurs frequently with the expression of aggression, anxiety itself instigates aggression on the basis of classical conditioning. As a stimulus (apart from its function as an inhibitor) anxiety may be supposed to add instigation to the aggression drive and to facilitate it, since the more relevant stimuli are present the more likely it is that a response will occur. This reasoning leads to the following hypothesis: If the anxiety evoked by a person's aggressive drive is strong enough, it should inhibit the aggressive response; however, if the anxiety is not strong enough to inhibit the aggresive response, the anxiety should facilitate the aggressive response that occurs. In simpler terms, an aggressive response that occurs in the presence of anxiety should tend to be stronger than an aggressive response that occurs in its absence. For example, the more anxious a person is about hitting another person, the harder he should tend to hit him if he does overcome his inhibition against aggression.

The child has to learn to satisfy his aggressive drive, but he has to do it in such a way that it does not call forth aggression anxiety. He must learn to alter his original violent forms of aggression to kinds that do not elicit so much retaliation from his family. One way of doing this is to attenuate the expression of aggression so that the persons against whom it is directed do not perceive the behavior as aggressive. The child learns through experience that there are certain kinds of aggression that the parents do not punish as much as other kinds. Most children are more likely to be punished for aggression that is overtly physical than for nonphysical (e.g., verbal) aggression; they are more likely to be punished for kicking their mother than for wetting the bed or performing some other act that the parents cannot so certainly perceive as aggression. This observation suggests a first principle: the more a child is punished for aggression toward any person, the more attenuated his overt expression of aggression toward that person becomes. The more punishment, the more anxiety, and the more anxiety, the more attenuation.

What happens to the fantasy aggression of the child as a result of the punishment for overt aggression? When aggression is expressed in fantasy, the anxiety associated with aggression against particular objects in real life situations tends to transfer by generalization to the fantasy situation. However, anxiety aroused through such a process of generalization tends to be weaker; hence, in the fantasy situation the aggressive response is more likely to occur than in the reality situation and any accompanying anxiety would tend to facilitate the response. Under these circumstances fantasy aggression is more rewarding for him the more violent it becomes. This reasoning rests on the assumption that direct undisguised aggression most adequately reduces the aggressive drive. A second principle is thus suggested: The more punishment that an aggressive act incurs in real life, the more direct and overt and violent is the aggression that occurs in fantasy.

These two hypotheses derive some support from the results of a study by Hollenberg and Sperry (1951). In the first part of their study, the experimenter verbally punished one group of children and did not verbally punish another group for showing aggression during doll play. The condition of verbal punishment is more similar in this respect to everyday life than to fantasy, whereas the condition of no verbal punishment represents a "pure" fantasy situation. Under the condition of verbal punishment, consequently, anxiety should act as an inhibitor and the expressed aggression be more attenuated (less "intense" in Sperry's terms) in the verbally punished group than in the control group; the data tend to support this prediction. Since Hollenberg and Sperry did not present the raw data in their paper, it is impossible to determine if this difference is significant.

---

[1] The writer wishes to express his deepest appreciation to Robert R. Sears, without whose aid this study would have been impossible. The expenses of the study were borne by a grant from the Behavioral Sciences Division of the Ford Foundation.

In the second part of their research, a different set of *S*s was divided into four groups according to experiences at home: those who had (*a*) low punishment for aggression and low frustration, (*b*) low punishment and high frustration, (*c*) high punishment and low frustration, and (*d*) high punishment and high frustration. These groups were compared with respect to the percentage of fantasy aggression that was "intense." In this study, the experimenter did not punish aggressive responses; consequently, the doll play responses represent fantasy responses and anxiety does not tend to be strong enough to act as an inhibitor. To test this hypothesis with frustration held constant, Group *a* is compared with *c* and Group *b* with Group *d*. In both comparisons, the hypothesis is that the group with the high punishment for aggression (and therefore high aggression anxiety) should show less attenuated (more "intense") aggression than the group with the low punishment; the data support this prediction, but again it is impossible to determine the significance of the differences.

If one assumes that more anxiety is aroused when aggression is directed toward adults than when aggression is directed toward other children, then it may be further hypothesized that in doll play in which aggression is not punished, the child's aggression toward the parent dolls tends to be more violent and direct (since it has the added instigation of anxiety) and that his aggression toward child dolls tends to be more attenuated. If aggression is considered to be more violent and less attenuated when it involves "bodily injury" than when it does not, one should expect a smaller proportion of aggressive responses involving bodily injury with child doll objects than with adult doll objects.

Although the same prediction is made about the doll agents of aggression, the reasoning behind the prediction is somewhat different than that underlying the prediction about doll objects. The objects of aggression in doll play represent the persons toward whom the child would like to direct aggression in real life; the agents of aggression in doll play presumably refer to the persons that the child would like to name as being aggressive. The child's parents punish him and this is aggression. However, the parents also punish him for saying that they are aggressive. Consequently, in real life situations anxiety is aroused whenever the child attributes aggressive acts to the parents. The expectation is therefore that a greater proportion of aggressive responses in doll play should be of the nonbodily injury type with child doll agents than with adult doll agents.

SOME RELEVANT DATA

The data to test this hypothesis were taken from a study previously reported by Pauline Sears (1951) and by Sears, Pintler, and Sears (1946). In this study, 150 children, 75 boys and 75 girls, 25 of each sex in the age groups three, four, and five years were given two 20-min. sessions of doll play with a standard doll family. The children were allowed complete freedom of response, and the experimenter did not punish aggressive behavior. Consequently, the child's responses in this study may be viewed as fantasy responses. Acts of aggression were recorded in detail and later scored in one of the following five categories: Type 1—the doll was used as an agent of bodily injury aggression; Type 2—the doll was used as an object of bodily injury aggression; Type 3—the doll was used as an agent of nonbodily injury aggression; Type 4—the doll was used as an object of nonbodily injury aggression; Type 5—all other forms of aggression.

To test the above hypothesis with the agents of aggression, an index of attenuation was calculated for each of the five dolls in each session by the formula: Type 3 ÷ (Type 1 + Type 3). (Children who did not show aggression of either type were omitted in computing the index.) Substituting Type 2 for Type 1 and Type 4 for Type 3 the same formula was used for objects of aggression. The higher these percentage values, the more relatively attenuated was the total aggression.

The significance of these differences was tested by calculating the CRs of the differences between the summed percentages of attenuated aggression for the two parent dolls with those for the three child dolls. These comparisons are shown in Table

TABLE 1

RELIABILITY OF DIFFERENCES BETWEEN SUMMED PERCENTAGE OF ATTENUATED AGGRESSION

	Agents						Objects					
	Session I			Session II			Session I			Session II		
	Parent Dolls		Child Dolls	Parent Dolls		Child Dolls	Parent Dolls		Child Dolls	Parent Dolls		Child Dolls
Boys	.54		.63	.43		.65	.28		.30	.20		.23
CR		.88			2.16			.22			.40	
P		NS			.03			NS			NS	
Girls	.69		.83	.54		.78	.37		.52	.24		.44
CR		1.10			2.31			1.34			1.97	
P		NS			.02			NS			.05	
Total[a]	CR = 3.10 *P* = .001						CR = 6.67 *P* = .0001					

[a] Represents the sum of males and females for both sessions.

1; two-tailed tests were used in computing the *P* values. The totals for agents and objects refer to the sums over both sexes and sessions. The order of events was as follows: the agents were tested for attenuation, the present theoretical explanation of the results was developed, and the prediction based on it then tested on the objects.

These results support the hypothesis of this paper fairly well, considering the crudeness of the measure of attenuation used. In fantasy as compared with real life, anxiety functions more as a stimulus than as an inhibitor.

REFERENCES

HOLLENBERG, E., & SPERRY, J. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32–43.

SEARS, PAULINE S. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**(6, Whole No. 323).

SEARS, R., PINTLER, M., & SEARS, PAULINE S. Effect of father separation on preschool children's doll play aggression. *Child Developm.*, 1946, **17**, 219–243.

(Received October 3, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

INDIVIDUAL DIFFERENCES IN REACTION TO FAILURE-INDUCED STRESS[1]

SAMUEL C. FULKERSON[2]
USAF School of Aviation Medicine

IN THE PAST a number of attempts have been made to find personality correlates with behavior under stress. Lazarus, Deese, and Osler (1952), reviewing the work up to 1951, indicated that scarcely any results of practical import had been found and emphasized the possibility of individual differences as a primary determinant of the stress reaction.

Since this review, several studies have appeared relating performance under stress to Janet's hysteria-psychasthenia personality typology. For instance, Eriksen (1954b) found that *S*s with higher scores on the Hysteria (*Hy*) scale of the Minnesota Multiphasic Personality Inventory showed greater stimulus generalization under stress than *S*s with higher scores on the Psychasthenia (*Pt*) scale. Deese, Lazarus, and Keenan (1953) reported results which suggest that learning under stress may be related to these same MMPI scores.

---

[1] This study is part of a project on the adaptability screening of flying personnel. Appreciation is expressed to Saul B. Sells and David K. Trites for their suggestions and criticisms, and to Wayne H. Holtzman for making available the data from the ROTC sample.

[2] Now at Western Psychiatric Institute, University of Pittsburgh Medical School.

The present study, conceived within the same frame of reference, is an attempt to relate scores on the *Hy* and *Pt* scales of the MMPI to performance under stress where the experimental task involves a kind of behavior which has been said to be important in distinguishing the hysteric personality type from the psychasthenic. The task was the McKinney Reporting Test (McKinney, Strother, Hines, & Allee, 1951), which introduces at its midpoint a psychological stress condition designed to elicit the threat of failure. The test requires *S* to count a number of small details in an accurate manner.

Insofar as the stress conditions arouse habitual defense reactions, the tendency for the hysteric to be careless about details and for the psychasthenic to be meticulous about them should be accentuated. The first hypothesis, then, was that psychasthenics would become more accurate than usual under stress and hysterics less accurate.

It also seemed likely that defense reactions would be stronger and more consistent for a poorly adjusted person than for a well adjusted person. Therefore, the second hypothesis was that the postulated differences in accuracy of performance between hysterics and psychasthenics would be greater the more poorly adjusted the individual.

METHOD

Subjects

For several years the USAF School of Aviation Medicine has supported a large scale research program on pilot selection. From this program four samples were available which had been administered the McKinney Reporting Test and the MMPI. Sample A consisted of 475 aviation cadets who were tested during the primary phase of pilot training. Sample B consisted of 428 copilots, tested during a three-month course in Combat Crew Training. Sample C consisted of 325 aircraft commanders, also tested during the course in Combat Crew Training. Sample D contained 118 ROTC officer cadets at the University of Texas.

Adjustment ratings of different kinds had been made within each of these samples. In Samples A, B, and C, adjustment was estimated primarily on the basis of a psychological examination, including an interview by a clinical psychologist, and the results from several projective tests. (Neither the MMPI nor the McKinney test were available to the raters.) In addition, training records, including medical records, demerits, academic grades, buddy ratings, and instructor ratings were also used to decide the adjustment of the trainee.

On the basis of these measures, two extreme groups were selected within each sample, one consisting of cases who showed clear-cut evidence of inability to adapt to the training program and who were also given a low adjustment rating by the psychologist. The other group contained individuals who had outstanding records during training as well as being rated high on adjustment.

From Sample A, 100 low adjustment and 100 high adjustment cases were selected to form Group A. Of Sample B, 93 low and 97 high adjustment cases formed Group B. From Sample C, 94 low and 97 high adjustment cases were selected for Group C. In this study the analysis was based on these groups rather than the total samples.

Sample D was treated differently because of its small size. Adjustment was measured by buddy ratings alone. Each *S* was asked to rank order all the men in his ROTC unit whom he knew well on a number of variables, including "personal adjustment." An average "personal adjustment" rating was computed on each individual and used as the adjustment measure. These ratings were then divided at the median, with the 58 cases below the median designated as low adjustment *S*s and the 60 above it as high adjustment *S*s.

Measures

The McKinney test consists of three parts: a prestress period, in which *S* can work at his own speed; a stress period, in which he is pressured to work beyond his capacity under the threat of failure; and a final poststress period, which returns to the conditions of the prestress period. Only measures from the prestress and stress periods were used in this study. The items on the test consist of diagrams containing small circles and lines, and the task is to count the number of these circles and dashes in each diagram. For the prestress and stress periods, the number of items attempted and the number of items correct were calculated. For each period, an accuracy score was computed by dividing the number of items correct by the number of items attempted. This gave the percentage of items correct. These percentage scores were transformed to standard scores.

In order to test the first hypothesis, that the threat of failure arouses the defense mechanisms typical for the hysteric or psychasthenic, a difference score was obtained between the accuracy scores by subtracting the score for the stress period from that for the prestress period. A constant was added to get rid of negative numbers. The *Hy* and *Pt* scales of the MMPI were then correlated with the McKinney difference scores.

Earlier studies (Fulkerson, 1957; Fulkerson, Freud, & Raynor, 1958) revealed that among pilots and pilot trainees, the *K* score was positively correlated with success in training, and its use as a "suppressor variable" decreased the discriminability of the scales. Therefore, the raw *Pt* scores, uncorrected by *K*, were converted into *T* scores. The *T* values used to convert *Pt* and *Hy* to standard scores were based on pilot norms, because the original norms were known (Fulkerson et al., 1958) to be inappropriate for a pilot sample.

RESULTS

The Pearson product-moment correlation coefficients between the McKinney and MMPI scores are shown in Table 1. An average of the coefficients for the four groups was obtained by the method outlined by Edwards (1950, p. 133). The correlations involving *Pt* were not significantly different from zero. The *Hy* score correlated significantly with the McKinney score in the predicted direction for the pooled low adjustment *S*s, with those having high *Hy* scores tending to shift toward greater inaccuracy, relative to the rest of the group, on being exposed to the stress period on the McKinney test.

The second hypothesis predicted stronger relationships for the low adjustment *S*s. Again, the correlations involving *Hy* are in line with this hypothesis.

The difference score derived from the McKinney test was designed to partial out the effect of initial differences in accuracy level from the accuracy score obtained on the stress period. It can be asked whether this gross correction did in fact improve the strength of the hypothesized relationships. Table 2 presents the correlations, averaged over the four groups, for the prestress and stress period accuracy scores on the McKinney test. The correlations for the prestress period are uniformly insignificant, and those involving the stress period are similar to those based on the difference score, though not consistently in the predicted direction.

TABLE 1

CORRELATIONS BETWEEN SHIFT SCORE ON THE McKINNEY REPORTING TEST AND THE MMPI SCALES OF *HY* AND *PT*

Group	Low Adjustment Subjects			High Adjustment Subjects		
	Hy	*Pt*	*N*	*Hy*	*Pt*	*N*
A	.17	−.03	100	−.22*	.10	100
B	.16	.04	93	−.05	−.19	97
C	.09	−.10	94	.09	−.13	97
D	.26*	.06	58	−.27*	.11	60
Average *r*	.16**	−.02		−.10	−.04	

* Significant at .01–.05 level.

** Significant at less than .01 level.

TABLE 2

AVERAGE CORRELATIONS BETWEEN THE ACCURACY SCORES ON THE PRESTRESS AND STRESS PERIODS OF THE McKINNEY REPORTING TEST AND THE MMPI SCALES OF *HY* AND *PT*

McKinney Test	Low Adjustment Subjects		High Adjustment Subjects	
	Hy	*Pt*	*Hy*	*Pt*
Prestress period	.03	−.04	.07	−.07
Stress period	−.13*	−.03	.17**	−.02

* Significant at .01–.05 level.

** Significant at less than .01 level.

DISCUSSION

The results of this study can hardly be said to demonstrate striking individual differences. They suggest that unless quite large samples are used, which contain a high proportion of poorly adjusted *S*s, the hypothesized relationship between performance under stress and the hysteria-psychasthenia dimension is not likely to appear. Even taking these factors into account, the personality dimension, at least as measured by the *Hy* scale, identified only 3% of the variance involved in the performance under stress. Of course *Hy* is not necessarily an accurate measure of the personality dimension, nor the McKinney test necessarily one that educes a powerful stress reaction. Another point to keep in mind is the homogeneity of these samples on several variables. All *S*s were relatively young, the ROTC student officers and the aviation cadets being primarily in their early twenties, the aircraft commanders and copilots in their late twenties and early thirties; all were above average in intelligence; there was homogeneity of vocational interests among the military samples; and, at least for the military samples, there was a strong set to do well on these tests, in all probability, since these *S*s had no reason to suppose the results might not affect their careers in some way. In a more heterogeneous sample, individual differences in reaction to stress might be clearer.

The results also indicate that *Hy* and *Pt* are not measures of a single bipolar trait. Eriksen (1954a) has reported similar findings and concluded that *Hy* is measuring the hysteric-psychasthenia dimension, whereas *Pt* seems to measure neuroticism or ego strength. In the present study, the over-all correlation between *Hy* and *Pt* was −.09. This correlation, however, was a function of adjustment level. For the low adjustment cases, the correlation was .06; for the high adjustment cases, it was

−.24. This latter value is significant beyond the .01 level.

These differences, as well as those shown in Table 1, indicate that adjustment level is a variable which must be controlled along with such standard variables as age and intelligence in personality studies.

SUMMARY

The McKinney Reporting Test and the MMPI were administered to samples of aviation cadets, aircraft commanders, copilots, and ROTC student officers. Two criterion groups representing the extremes in adjustment were selected within each of the military samples to form high and low adjustment groups. The ROTC officers were divided at the median into upper and lower adjustment groups.

There was an over-all significant tendency among the low adjustment cases for those who shifted toward more accurate performance on the McKinney stress period to answer on the *Hy* scale of the MMPI like Janet's psychasthenic type, and for those who became more inaccurate to answer like the hysteric. This tendency did not hold for high adjustment cases. The McKinney test did not correlate with the *Pt* scale.

REFERENCES

DEESE, H., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50. (a)

ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–564. (b)

FULKERSON, S. C. Adaptability screening of flying personnel: Research on the Minnesota Multiphasic Personality Inventory. *USAF Sch. Aviat. Med., Rep.*, 1957, Proj. No. 57–106.

FULKERSON, S. C., FREUD, S. L., & RAYNOR, G. H. The use of the MMPI in the psychological evaluation of pilots. *J. aviat. Med.*, 1958, **29**, 122–129.

LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

MCKINNEY, F., STROTHER, G. B., HINES, R. R., & ALLEE, R. A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.

(Received October 27, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

A CROSS-CULTURAL STUDY OF STRUCTURED GROUPS WITH UNPOPULAR CENTRAL MEMBERS

A. I. MOHANNA
Al-Azhar University, Cairo

AND MICHAEL ARGYLE
University of Oxford

THIS paper reports an experiment on the performance of small social groups with restricted channels of communication. The method used by Leavitt (1951) was followed, and the following additional variables were manipulated: (*a*) the central member of the wheel groups was selected sociometrically to be popular or unpopular (the validity of this manipulation being confirmed by a postexperimental questionnaire), (*b*) the experiment was repeated in Oxford and in Cairo, with all other conditions held as constant as possible. Altogether 80 groups were used, involving 400 schoolboys aged 13½–15½.

Results

Wheel vs. circle patterns. Leavitt's result was confirmed that the wheel pattern is superior in time taken, number of messages required, and errors made (see Table 1).

Wheel pattern with popular vs. unpopular central members. As can be seen in Figures 1 and 2, the unpopular wheel groups settle down to a level of performance better than the circle groups, and worse than the popular wheel groups, in time taken and messages used, and "learn" faster than the other two kinds of groups.

Oxford vs. Cairo. Table 1 indicates that there

TABLE 1
WHEEL v. CIRCLE PATTERN—RESULTS FOR TRIALS 17–24

		Oxford		Cairo	
Time taken (secs.)	Wheel	43.1		32.5	
	Circle	103.1	($p < .01$)	83.4	($p < .001$)
Number of messages	Wheel	10.0		8.1	
	Circle	13.9	(n/s)	16.0	($p < .001$)
Number of errors	Wheel	0.78		0.34	
	Circle	1.31	(n/s)	0.92	($p < .01$)

Note.—n = 8 for each condition. The t test was used throughout, with a one-tailed distribution.

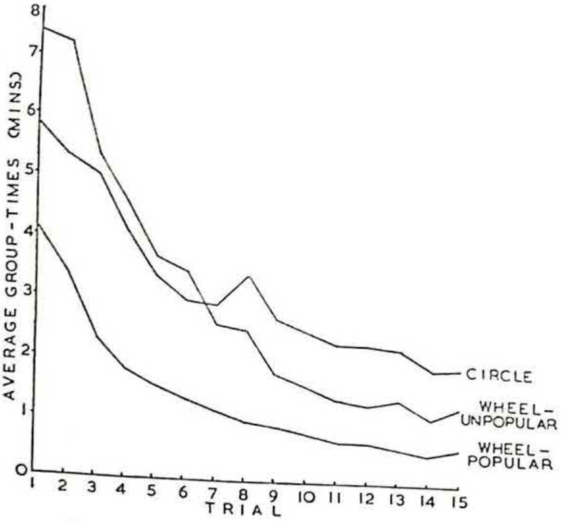


FIG. 1. AVERAGE GROUP TIMES PER TRIAL (16 GROUPS ON EACH CONDITION)

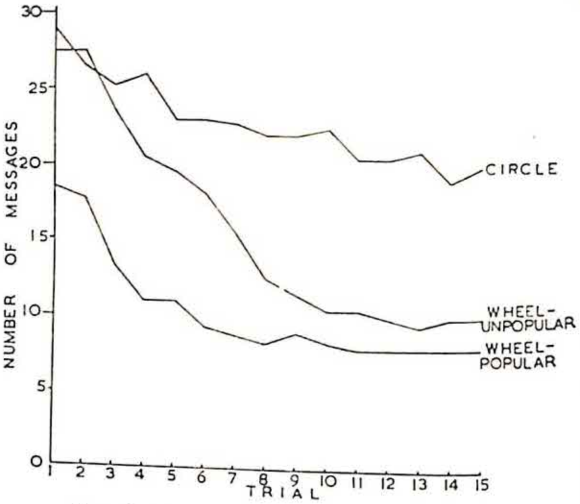


FIG. 2. AVERAGE MESSAGES PER TRIAL (16 GROUPS IN EACH CONDITION)

are only slight (and insignificant) differences in absolute efficiency between the Oxford and the Cairo groups. Satisfaction was greater in popular wheel groups, followed by circle groups and unpopular wheel groups; this result was found in Cairo only, but the generally higher level of satisfaction of the Oxford *S*s may have concealed any differences. Otherwise the results in Oxford and Cairo were closely similar.

Discussion

The inefficiency of unpopular compared with popular wheel groups may be due to the same causes as the low output of noncohesive groups on cooperative group tasks—that there is an additional incentive for work (Argyle, 1957, pp. 126–8). The more rapid learning in unpopular wheel groups could be due to a short-term version of "interaction leads to liking"; that is, the groups are extremely slow at first, but greater speed becomes possible as the initial hostility is *partly* reduced by interaction.

REFERENCES

ARGYLE, M. *The scientific study of social behaviour.* London: Methuen, 1957.

LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38–50.

(Received September 4, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

RECALLED DREAM THEMES AND PERSONALITY[1]

JOSEPH F. RYCHLAK
Washington State University

ALTHOUGH ACCEPTED as legitimate data for personality analysis by most clinicians, dreams have only occasionally been studied systematically in relation to personality variables. The present study of children's dreams was designed to test the hypothesis that persons dream in a pat-

---

[1] The author would like to thank the following people for their help with the study: J. G. Wynn Jr., Mary C. Witt, and J. H. Diamond of University School, Florida State University; C. Moos of the Psychology Department, Florida State University; Alice O. Thomas of Pullman, Washington.

terned fashion and that such uniformities can be related to objective test findings, and possibly to sociometric ratings. The fact that these dreams are "recalled" is given at the outset. Paralleling the clinical procedure of weekly sessions with a client, the study concerned dreams recalled by *S*s over the past week.

METHOD

Dream Measurement

Each dream collected was given a single scoring, according to its major theme. A major theme was seen as reflecting the manifest point, or story line of the dream, which usually was easily identified because *S*s reported only the core of their dreams. In longer and more complex reports the theme scored as major usually revolved about the final event of the dream story.

Dreams were scored for three major themes: Affiliative, Reward, and Tension. A method of scoring was devised on the basis of previous research findings (Alexander & Wilson, 1935; Blanchard, 1926; Jersild, Markey & Jersild, 1933; Witty & Kopel, 1939) and preliminary exploration. A brief description of each category with examples taken from *S*s' reports follows:

Affiliative: Dreams having relaxed or pleasurable interpersonal relations with other human beings as their major themes, regardless of the latter's relationship to *S*; e.g., relatives, peers, same or opposite sex, etc. There must be some form of interaction with the person, and not simply a recollection of seeing him in the dream.

Examples:

I dreamed that my brother and I went swimming.

I dreamed that a boy in our room took every girl out to a show.

I dreamed that me and Stevie played catch.

Reward: Dreams with major themes involving a positive, pleasurable connotation *other than* that derived from strictly interpersonal contact. Reward themes include individual achievement and recognition, the receiving of gifts, and any phantasy-like tale of miraculous happenings.

Examples:

I dreamed I got a new watch.

I dreamed I scored 25 points against Panama City.

I was on Arthur Godfrey's Talent Scouts.

Tension: Dreams reflecting anxiety, frustration, or hostility in their major themes.

Examples:

I dreamed that Frankenstein was after me.

I was riding my bike and the wheel came off.

I dreamed some friends of mine were killed in an auto accident.

Less than 5% of the dreams were unscorable. A reliability check was performed on 32 randomly chosen dreams. The *E* and a trained rater independently scored each dream with 94% agreement (agreements divided by total comparisons).

Personality Measurement

Cattell's Junior Personality Quiz (Cattell, Beloff, Flint, & Gruen, 1953) was used, a 12-factor test applicable to the age range of 10 to 16 years. It contains 144 items and takes approximately one hour to administer.

Popularity Measurement

Popularity was assessed sociometrically. Each *S* selected three children (either sex) with whom he would "like best to play any of the games that we know" and three children with whom he would "like to work or make something." The popularity score was the absolute number of choices an *S* received from his classmates.

Subjects

Two elementary school classes were studied, a fifth grade (17 females, 13 males) and an eighth grade (15 females, 14 males). The final sample of children who reported at least two dreams included 25 girls (13 fifth, 12 eighth graders) and 17 boys (7 fifth, 10 eighth graders). In general the children had middle-class backgrounds. Mean CAs were 10-9 for the fifth graders and 14-1 for the eighth graders. Mean grade-level scores for Total Language Ability on The California Achievement Test (Tiegs & Clark, 1951) were as follows: fifth graders 5.99, eighth graders 8.87.

Procedure

Dream themes were collected in writing on the same day of each week over a three-month period. When more than one theme was recalled it was accepted. If a child could not recall having a dream over the past week he was asked to indicate this in writing. If he was absent on the day of collection he reported his dream on the first day of his return. Emphasis was repeatedly placed on honesty in reporting, and reassurance was given that not all children have dreams and many who do sometimes forget them. After 12 weeks each *S* was requested to make up a "daydream" in class.

Data obtained in this way have obvious limitations. A child might make up a dream story to please authority figures. *S*s with poor verbal skills may neglect reporting dreams because of the effort it takes to record them. To check on these factors certain measures were taken.

First, the daydreams were used as a form of

TABLE 1
DREAM PRODUCTIVITY AND MAJOR THEMES
($N = 59$)

Sample Breakdown	Dream Incidence		Percentage of n reporting one or more themes of		
	Median	Med. Devia.	Affiliative	Reward	Tension
All Boys ($n = 27$)	2.20	1.48	26*	59	67
All Girls ($n = 32$)	2.90	2.33	56	44	72
5th Graders ($n = 30$)	2.33	2.22	40	57	67
8th Graders ($n = 29$)	2.62	1.83	45	45	73

* Significantly different than comparable percentage for girls: $P = .02$.

check on the presumed legitimate dream data. It was found that these imagined dreams were consistently longer, generally more coherent and logical in their themes, and much more likely to have well defined, happy endings. On the basis of such principles a trained rater checked "actual" or "daydream" following the reading of 10 dreams of each type randomly selected, typed on plain paper and shuffled before rating. He was able to discriminate correctly in 75% of the cases ($P = .042$).

Secondly, comparisons were made between the number of dreams submitted by *S*s and their scores on two indices presumably related to verbal skills: intelligence as measured by the personality test (Factor 12, Cattell et al., 1953) and Total Language Scores of The California Achievement Test (Tiegs & Clark, 1951). These chi squares failed to reach even the .10 level of significance.

RESULTS

Table 1 reports the median number of dreams collected, and the percentage of *S*s producing at least one theme in each of the three dream categories. Thus, 26% of the boys and 56% of the girls reported at least one Affiliative theme over the 12 weeks of the study, the only significant difference in the table. Considering only those *S*s who recalled two or more dreams (to whom subsequent analysis was restricted by prior decision) the median values became: girls ($n = 25$), Median 3.44, Median Deviation 2.30; boys ($n = 17$), Median 3.20, Median Deviation 1.15. In order to assess possible influences of pubescence on major themes, the Affiliative dreams were rescored on the basis of obviously romantic content (e.g., "I dreamed that Johnnie kissed me"); 25% of the fifth graders and 27% of the eighth graders reported from one to four heterosexual themes.

The *S*s who produced two or more dreams were ranked with respect to the proportion of reported dreams classified under each theme. In the ensuing analysis, the sample was dichotomized at the median on each measure under consideration, and null hypotheses of independence evaluated by chi square.

Dreamers falling above the median on Affiliative themes fell below the median on sociometric play choices (chi square, 4.71, $P = .04$). Reward dreamers tended, but not significantly so, to fall above the median in play choices (chi square, 2.24, $P = .15$). On the work group choices the Affiliative dreamers once again fell below the median in popularity (4.68, $P = .04$) and the Reward dreamers showed a comparable tendency to popularity (2.40, $P = .13$). Tension dreamers failed to show significant consistencies of choice in either group (chi squares of 1.61 and .088). When the choices for both work and play groups are considered as a single sociometric score, the chi squares increase as follows: Affiliative themes and low popularity 7.78 ($p < .01$); Reward themes and tendency to popularity 2.66 ($P = .11$).

On the personality variables, seven comparisons were accepted as reflecting a significant difference. Affiliative themes bore a relationship to the following three personality variables: Emotional Sensitivity (9.54, $P < .01$), Schizothymia (4.71, $P = .04$), and Socialized Morale (4.71, $P = .04$). Reward dreamers fell within the personality variables of Cyclothymia (3.38, $P = .07$) and Independent Dominance (12.30, $P < .001$). Tension dreamers scored high on the variables of Withdrawn Schizothymia (4.87, $P = .04$) and Lowered Intelligence (4.71, $P = .04$).

DISCUSSION

The hypothesis under consideration has received some verification. Manifest dream themes have been shown to discriminate fairly well between *S*s and to relate to personality and sociometric measurements. Two of the factors bearing a significant relationship to dreams are among the least reliable of the Personality Quiz (split-half r): Cyclothymia vs. Schizothymia (.31) and Independent Dominance (.30). Interpretations made on the basis of these variables must be considered tentative. On the other hand, the two most reliable factors of the test also entered into the findings: Emotional Sensitivity (.83) and Intelligence (.78).

In spite of the gross classification employed in this study, a review of the findings in conjunction with the definitions of the measured personality factors (Cattell et al., 1953, pp. 8–11) suggests certain common personality trends among dreamer types. Passive-conformity, inhibition, and a preference for close interpersonal relations seem to typify the Affiliative dreamer. This motivational pattern may be viewed as either cause or result of the relatively low popularity disclosed by the sociometric data. The Reward dreamer seems to be either outgoing and affable or fairly dominant and achievement-oriented. Needs reflected in Reward themes are not inconsistent with these behaviors. *S*s having many Tension dreams scored highest on one of the factors reflecting neurotic trends, as might be expected from clinical experience. Their lower standing on intelligence may reflect a relationship between intelligence and frustration in school tasks.

REFERENCES

ALEXANDER, F., & WILSON, G. W. Quantitative dream studies. A methodological attempt at quantitative evaluation of psychoanalytical material. *Psychoanal. Quart.*, 1935, **4**, 371–407.

BLANCHARD, P. A study of subject matter and motivation of children's dreams. *J. abnorm. soc. Psychol.*, 1926, **21**, 24–37.

CATTELL, R. B., BELOFF, J., FLINT, D., & GRUEN, W. *Handbook for the junior personality quiz.* Champaign, Ill.: Inst. for Person. & Ability Testing, 1953.

JERSILD, A. T., MARKEY, F. V., & JERSILD, C. L. Children's fears, dreams, wishes, daydreams, likes, dislikes, pleasant and unpleasant memories. *Child Development Monographs*, 1933, No. 12.

TIEGS, E. W., & CLARK, W. W. *Manual of directions for the primary California achievement test battery.* Los Angeles: California Test Bur., 1951.

WITTY, P. A., & KOPEL, D. The dreams and wishes of elementary school children. *J. educ. Psychol.*, 1939, **30**, 199–205.

(Received November 26, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SCHIZOPHRENIC-LIKE RESPONSES IN NORMAL SUBJECTS UNDER TIME PRESSURE[1]

GEORGE USDANSKY AND LOREN J. CHAPMAN[2]
University of Chicago

A NUMBER OF THEORISTS and investigators have discussed the formal similarities of the thinking of schizophrenics to that of normal *S*s under special conditions. Soon after Freud described primary process thinking in dreams, Jung pointed out similarities between dream thinking and the thinking of schizophrenics. Jung also reported similarities on his word association test between schizophrenics and normals in states of drowsiness or reduced attention. Other conditions that have since been reported as producing schizophrenic-like thinking in normals include organic brain damage, drug induced states, oxygen deprivation, personalized distraction, fatigue, and time pressure. Cameron and Magaret (1951) add several more to the list, including interruption of ongoing activity, environmental changes, preoccupation, and emotional excitement.

It appears easy to fit atypical normal thinking into categories that have been used to describe schizophrenic thought, perhaps in part because of the vagueness of the categories. In drawing such comparisons—or in drawing any conclusions about thinking pathology—it is important to specify exactly the nature of the thinking deviation. The present investigation of a condition under which normal *S*s show an increased tendency to make a schizophrenic kind of error uses a method that makes possible precise specification of the nature of the error. In place of using the traditional method of observation followed by after-the-fact classification, *E* administered a concept sorting task that has built into it opportunities for making specifically associative errors. The task has previously been shown to distinguish a schizophrenic and a normal group by the number of incorrect responses that are associative in nature (Chapman, 1958). Such errors appear on the basis of clinical evidence to be important in the thinking deviation of schizophrenics (Arieti, 1955; Bleuler, 1950; Jung, 1936; Kraepelin, 1919).

The special condition investigated here was that of speeding up responses by instructing *S* to hurry. It was hypothesized that speeding up the responses of normals would produce an increase in associative errors.

METHOD

Procedure. The *S* was asked to place a sorting card bearing the name of a common object, e.g.,

---

[1] This study was supported in part by a research grant (M-1416) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Now at the University of Kentucky.

TABLE 1

EXAMPLES OF TEST ITEMS

Sorting Card Word	Guide Card Words[a]		
gold	typewriter (I)	steel (C)	fish (A)
hat	head (A)	dress (C)	garage (I)
raincoat	pencil (I)	water (A)	awning (C)
lip	stick (A)	ear (C)	chair (I)
dragon	rope (I)	trout (C)	wagon (A)
face	lace (A)	weed (I)	knee (C)

[a] (C) identifies a word as the correct conceptual choice, (A) the associative distracter, and (I) the irrelevant. See text for further explanation.

"gold," with that one of three guide card words which "could be called by the same name or term" as the sorting card word. In the "gold" example, the three guide card words are: typewriter, steel, and fish. The correct answer is "steel" because both it and "gold" can be called "metal." "Fish" is an associative distracter on the basis of commonly experienced contiguity of words. "Typewriter," on the other hand, is a word intended to have neither conceptual closeness nor associative connection with the sorting word. It is called here an "irrelevant" word. Two other types of associative distracters were used, associations based on contiguity of objects in common experience, e.g., hat–head, raincoat–water, and those based on rhymes, e.g., dragon–wagon, face–lace. Table 1 lists six of the items. Increased speed of sorting could be expected to result in an increase in errors of all types. But errors other than in response to the associative distracter words should, by chance, be equally divided among placements with the associative guide card word and the irrelevant one. These errors contribute, in proportion to their number, to raising the total of associative errors. Therefore, a measure of the effect of associative pull alone is a score of associative placements minus the irrelevant placements. This association-minus-irrelevant score should yield a comparatively pure measure of response to the associative connections. Since an earlier study showed that this error score is responsible for much of the inferiority of schizophrenics to normals on this task, an increase of this score in normals under time pressure may be described as an increase in schizophrenic-like errors.

Ninety test items like those described were used. The 90 were divided into two blocks of 45, with each of the three bases of association equally represented in each block. Each *S* sorted one block of cards under the fast condition and the other under the slow. The two blocks appeared equally often in the two conditions. To control for the differential effects of practice and fatigue, each of the two experimental conditions was equally often presented first and second.

Instructions. The *S* was instructed: "For each card in your hand, you decide which word up here mentions something that can be called by the same name or term, and you place your card in the correct tray." This instruction was followed by six examples accompanied by further explanation. The *S* who was to sort under the slow condition first was then told: "Now go ahead on your own with the rest of the cards in the pile before you. Your speed is not important. I want you to take your time and do your best on each card." After completing the first block of cards, he was given the alternate block with instructions stressing that he was to sort these at his topmost speed, that the interest on this block of cards was in "just how fast you can get through these cards while still trying to get them right." Instructions were substantially the same, but in reverse order, for those *S*s who sorted under the fast conditions first.

Apparatus. Sorting cards were 3 × 5. Each guide card, 2 × 11, had its three guide card words printed on a single line. The trays were of cardboard, ⅝ deep and 7 square. The guide cards were placed on the table so that they were directly in *S*'s line of vision, with a tray in front of each guide card word. *E*, holding the guide cards on the table, sat opposite *S* and removed the top card just before *S* completed his placement in one of the trays. Thus there was continually a new guide card confronting *S*.

Subjects. The *S*s were 28 junior college and teacher's college male students ranging from freshmen to seniors and in age from 18 to 41 years, with a mean age of 24.4 years. Three additional *S*s were tested but were not included in the analysis of results because the differences between their "fast" and "slow" sorting times were atypically slight, indicating that the *S*s were not following the instructions. These atypical differences (20, 17, and 4 sec.) were all more than 2 *SD* units from the group mean difference.

RESULTS

The *S*s responded to the instructions for fast and slow sorting, as shown by a mean sorting time of 114 sec. under the fast condition and 220 sec. under the slow.

Table 2 lists the means and medians for the various error scores under both fast and slow conditions, showing scores separately for the three bases of association as well as for combined association and combined irrelevant scores. The asso-

TABLE 2

ERROR SCORES UNDER THE FAST AND SLOW CONDITIONS

Score and Associative Distracter	Fast Condition		Slow Condition		Difference (Fast minus Slow)		
	M	*Mdn*	*M*	*Mdn*	*M*	*Mdn*	*CR*
Association choices							
Rhyme	1.93	1.10	.68	.41			
Word	3.21	2.30	1.93	1.20			
Object	4.89	4.50	3.43	1.25			
Combined	10.04	9.00	6.04	2.50	4.00	3.50	3.12
Irrelevant choices							
Rhyme	.75	.50	.14	.28			
Word	.71	.60	.07	.27			
Object	.39	.35	.04	.26			
Combined	1.86	1.39	.25	.30	1.61	1.36	3.90
Association minus Irrelevant							
Rhyme	1.18	.30	.54	.25	.64	.13	.98
Word	2.50	1.83	1.86	1.17	.64	.83	1.74
Object	4.50	4.00	3.39	1.25	1.11	1.00	1.49
Combined	8.18	6.83	5.79	2.50	2.39	2.00	2.25

ciative basis of any given irrelevant error is determined by the type of association error possible on that same item.

Because of the strongly skewed distribution of both associative and irrelevant error scores, a Wilcoxon signed rank test was used to test the significance of the difference between various kinds of errors. All values were computed using a two-tailed test. The associative error scores exceeded the irrelevant error scores ($p < .01$) for both the fast and slow conditions.

The third section of Table 2, headed Fast minus Slow, contains the means, medians, and critical ratios for the differences between conditions. Taking the three bases of association together, we see that, as expected, both the associative and the irrelevant error scores increased under the fast condition ($p < .01$). Of greatest interest, however, is the increase under the fast condition of the association-minus-irrelevant score ($p < .03$). This score, it will be recalled, is the purer measure of associative intrusion. We conclude that normal performance under time pressure was significantly impaired due to an increase in associative errors.

Table 2 further shows that all three bases of associative distraction contributed to this significant difference, between conditions, of the total association-minus-irrelevant score. Taken separately, however, none of the three attained significance.

DISCUSSION

This experiment has shown that certain errors made by normals under time pressure are like the errors of schizophrenics. Chapman (1958) has recently given evidence that the occurrence of schizophrenic-like errors in normals is not dependent on "special conditions," but that such errors, usually described as aspects of the schizophrenic thinking disorder, are also found in undisturbed normal functioning, only with less frequency. The present experiment provides further evidence for this view in that even under the slow condition *S*s made associative errors more often than irrelevant ones. Under the fast condition, however, the *S*s became more schizophrenic-like in their errors.

To speak of a "special condition" in an experiment of this kind may exaggerate the degree to which the independent variable departs from ordinary experience. Many, and perhaps all, special conditions that have been discussed as such are found to some extent in any state of normal functioning, e.g., fatigue, distraction, oxygen deprivation. Because these special conditions may each be viewed as a continuum, it may be possible to order given characteristics of schizophrenic performance in terms of the ease with which they can be elicited from normals under the "special conditions." Thus, while a given period without sleep might be necessary to produce susceptibility in normals to certain kinds of errors, e.g., associative errors, a greater period might be found necessary to produce others, e.g., neologistic combining of word units. Such a determination would express the availability of these errors to normal thinking and would be relevant to the question of the degree of schizophrenic disturbance represented by a given error. Moreover, it is possible that some kinds of schizophrenic errors are not made by normal *S*s. Evidence to support this view, or the opposite one mentioned at the beginning of this discussion, can be obtained only by inducing considerable errors in normals, as was done in the present study. Any such error-inducing procedure might for convenience be called a "special condition," even though it may differ from ordinary experience quantitatively rather than qualitatively.

The findings of the present experiment may be interpreted as supporting the microgenetic theory of thinking as formulated by Schilder (1951) and recently reviewed by Flavell and Draguns (1957). According to these theorists, abnormal thought processes such as are found in schizophrenics, brain damaged patients, "primitives," children, dreamers, and normals under special conditions are incomplete, early phases of development of individual thoughts which have emerged into consciousness as though they were the final stages of the thought development. Flavell and Draguns (1957) described these early phases of thought as resembling Freud's primary process thinking, and they, like Freud, include responding inappropriately to associative connections among its characteristics. They suggested that the microgenetic theory might be tested by an experiment in which responses are made under time pressure, because rapid response may be thought of as fostering the emergence of incomplete thoughts. Soon after this, Flavell, Draguns, Feinberg, and Budin (1958) published such an experiment, which had some similarity to the present study. Using a word association test, they compared the differences between schizophrenic and normal responses with the differences between responses of normals with and without time pressure. They found some significant differences in the predicted direction, and they interpreted their findings in the context of the microgenetic formulation. Since the present study also demonstrated an increase in a kind of response attributed by Schilder (1951) to the early phases of thought, these results, too, might be interpreted as supporting the microgenetic theory. However, because the microgenetic theory is "a highly speculative view of a variety of cognitive phenomena" (Flavell & Draguns, 1957, p. 212), we prefer to limit our interpretation of findings to the empirical statement, that a specific kind of error which occurs more in schizophrenics than normals also occurs more in hurried responses. One can never be sure that he is observing an incomplete thought. What we do observe is that the errors made in schizophrenia and in various other nonnormal conditions appear to resemble one another. The independent variable of this study is just one of the many conditions that have been described as producing schizophrenic-like errors, and it has no logical primacy in accoutning for such similarities.

Flavell and Draguns (1957) based much of their discussion on these similarities. However, a great deal of research is needed to determine just what are the similarities and differences between the errors made under the various special conditions. Alleged similarities between disorders may diminish under closer scrutiny. For example, the similarity of thinking disorders of schizophrenics and cerebral organics, as described by Bolles and Goldstein (1938) and by Goldstein and Scheerer (1941) was found by McGaughran and Moran (1957) to be at least in part an artifact of the crudity of Goldstein's classification scheme.

SUMMARY

It was hypothesized that the conceptual performance of normal *S*s working under time pressure would deteriorate as a result of an increase in associative intrusions, and would thus more nearly resemble the performance of schizophrenics. Twenty-eight *S*s were given a conceptual card sorting task in which for each sorting choice there were three alternatives, one of which was the correct conceptual response, one an associative distracter, and one which was neither, called here the irrelevant response. Since errors due to sources other than response to the associative distracters should be equally divided among placements with the three guide cards, a score of associative placements minus irrelevant placements yields a comparatively pure measure of response to the associative intrusions as such. This score was significantly higher under the fast sorting condition than under the slow. Since the same measure had been previously found to be much higher for schizophrenics than for normals, we may conclude that increasing the speed of response produced an increase in a schizophrenic-like kind of error.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: Internat. Univer. Press, 1950.

BOLLES, MARJORIE M., & GOLDSTEIN, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42–65.

CAMERON, N., & MAGARET, ANN. *Behavior pathology.* Cambridge, Mass.: Riverside, 1951.

CHAPMAN, L. J. The intrusion of associative responses into schizophrenic conceptual performance. *J. abnorm. soc. Psychol.*, 1958, **56**, 374–379.

FLAVELL, J. H., & DRAGUNS, J. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54**, 197–217.

FLAVELL, J. H., DRAGUNS, J., FEINBERG, L. D., & BUDIN, W. A microgenetic approach to word association. *J. abnorm. soc. Psychol.*, 1958, **57**, 1–8.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior; an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

JUNG, C. G. The psychology of dementia praecox. *Nerv. ment. Dis. Monogr. Ser.*, 1936, No. 3.

KRAEPELIN, E. *Dementia praecox and paraphrenia.* Edinburgh: Livingstone, 1919.

MCGAUGHRAN, L. S., & MORAN, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *J. abnorm. soc. Psychol.*, 1957, **54**, 44–49.

SCHILDER, P. On the development of thoughts. In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951. Pp. 497–580.

(Received November 26, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

EPPS n ACHIEVEMENT AND VERBAL PAIRED-ASSOCIATES LEARNING

LEONARD WORELL[1]
State University of Iowa

IN RECENT YEARS three principal techniques have been proposed as measures of an achievement personality variable. The predictive utility of two of these, that developed by McClelland, Atkinson, Clark, and Lowell (1953) and the Iowa Picture Interpretation Test (IPIT) (Hedlund, 1953), has received considerable attention, while relatively little has appeared on the n Achievement scale in the Edwards Personal Preference Schedule (EPPS) (Edwards, 1954). The present study is concerned with an attempt to determine some of the theoretical and empirical implications of EPPS n Achievement.

Several recent studies indicate that the three techniques, though presumably assessing the same variable, have consistently low and insignificant interrelationships (McClelland, 1958, pp. 37–38; Melikian, 1958).[2] These findings suggest that one cannot readily anticipate relationships between n Achievement and other variables for any one measure on the basis of the results obtained with either of the other two instruments.

Rather than using the findings of previous studies on n Achievement as a basis for prediction, an alternative approach may be made through the analysis of the apparent properties associated with the EPPS achievement variable. Edwards' designation of achievement as a manifest need together with the verbal statements that he associates with the term seem to imply that the construct possesses motivational properties. In his description of n Achievement, Edwards (1954, p. 5) includes the following characteristics:

> To do one's best, to be successful, to accomplish tasks requiring skill and effort, to be a recognized authority, to accomplish something of great significance, to do a difficult job well, to solve difficult problems and puzzles, to be able to do things better than others, to write a great novel or play.

These statements imply at least the following: first, endorsement indicates that a strong desire for or a high value is placed on achievement, and, secondly, high scores are associated with *general* achievement rather than selected achievement situations, such as the academic, athletic, etc. Neither of these aspects, however, leads to clear-cut predictions as to the manner in which n Achievement is related to learning or performance. Furthermore, Edwards does not indicate what, if any, relationships may be expected to other situations.

In the absence of such considerations, it may be useful to examine the role of achievement as a motivational variable in learning situations in the light of two general classes of variables—associative and nonassociative. As Farber (1955) has suggested, this distinction leads to the classification of certain motivational variables as reflecting primarily steering or directional (associative) properties as compared with those representing essentially energizing or drive (nonassociative) properties. From an empirical standpoint, the presence of an associative variable is suggested by its tendency to facilitate a limited class of responses only, while a drive variable is indicated in part by its tendency to energize indiscriminately all response tendencies existing in a given situation.

In the present study, performance on two lists of verbal paired-associates learning materials is used to examine the role of n Achievement as a possible drive variable. These lists have been used previously with *S*s differentiated by an assumed measure of drive—the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953). Specifically, on a list comprised of paired nouns, highly anxious *S*s have not differed from nonanxious *S*s, while on a list of adjectives paired for high association value, highly anxious *S*s have demonstrated superior performance over nonanxious ones (Spence, Taylor, & Ketchel, 1956). If the EPPS n Achievement scale reflects the properties of a drive variable similar to those assumed for the MAS, differences between *S*s varying in n Achievement should parallel those obtained with the MAS in the two verbal learning situations.

METHOD

Subjects. The *S*s were 38 males and females from the introductory psychology course at the State University of Iowa. The group high in EPPS n Achievement consisted of 13 females, whose scores fell at or above the 79th percentile, and 6 males, with scores at or above the 74th percentile. The low n Achievement group was composed of 13 females, with scores at or below the 8th percentile, and 6 males, with scores at or below the 22nd percentile.

---

[1] Now at Oklahoma State University.

[2] The investigator has obtained a nonsignificant correlation of −.09 between EPPS and IPIT Achievement with a sample of 102 college *S*s.

TABLE 1
NOUN AND ADJECTIVE TEST LISTS

Noun Pairs		Adjective Pairs	
Stimulus	Response	Stimulus	Response
sand	heart	distant	remote
rose	jaw	wicked	corrupt
marsh	quart	insane	deranged
flute	barb	haughty	aloof
tile	palm	barren	fruitless
alms	sang	little	minute
ledge	dime	roving	errant
year	rack	mammoth	oversize
chalk	zoo	tranquil	placid
wool	tomb	empty	vacant
glass	yeast	playful	genial
iron	leaf	adept	skillful
kink	elm	complete	thorough
nail	vest	stubborn	hidebound
unit	oil		

Apparatus. Two lists of paired-associates learning materials were presented in Hull memory drums. Successive stimulus items of each list were exposed every 4 sec., with a 1.67-sec. anticipation interval. Three orders of presentation were used for each list with a 4-sec. rest interval between each order.

The first list (Table 1), consisting of 15 paired nouns, has been used previously (Spence et al., 1956) to acquaint *S* with the procedure and to establish minimal and maximal performance criteria. The second list (Table 1), 15 two-syllable adjectives obtained from Haagen's (1949) word list, had been constructed in another study (Spence et al., 1956) with the intent of maximizing the strength of association between the paired stimulus-response words. An attempt was also made in the construction of the list to minimize meaningful intralist associations and formal similarities among the stimulus-response pairs.

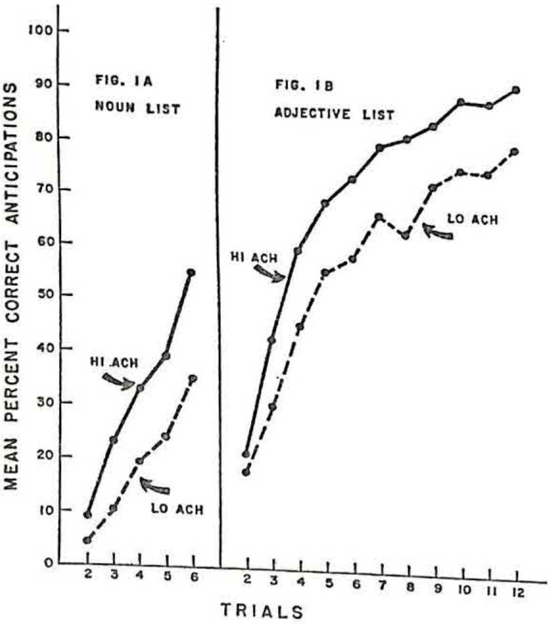


FIG. 1A AND 1B. Paired-associates learning as a function of n Achievement on noun and adjective lists.

PERFORMANCE ON ADJECTIVE LIST
TABLE 2

Group	*N*	Trials to Criterion		% Correct Anticipations	
		Mean	*SD*	Mean	*SD*
Hi-Ach	19	13.47	8.25	71.33	16.79
Lo-Ach	19	19.26	8.04	59.02	17.60

Procedure. Each *S*, participating individually, was exposed to six trials of the noun list. In the 2-min. rest period following exposure to the first list, *S* was moved to a seat before the screen containing the drum with the adjective list. Exposures to the adjective list were continued until *S* attained two successive errorless trials or had made 34 trials without reaching this criterion. A record was kept of correct anticipations and errors, the latter consisting of making either no response or an incorrect response.

In previous studies it has been typical to discard *S*s on the basis of their scores on the noun list, i.e., failure to make a single correct response or making 50 or more correct responses during a six-trial period. Since differences in verbal learning were considered meaningful for an interpretation of achievement, only the minimal performance criterion was invoked. No *S*, however, made fewer than one correct response.

RESULTS

Figures 1A and 1B contain the curves for the group high and low in EPPS n Achievement according to the mean percentage of correct anticipations by trial for the noun and adjective lists respectively. Both sets of curves rise continuously and the high n Achievement group is consistently superior to the low n Achievement group.

The mean number of correct anticipations on the noun list was 24.1 for the high n Achievement group and 14.1 for the low group. Analysis of variance indicated that the difference associated with achievement is significant beyond the .01 level ($F = 8.69$), while neither sex nor the sex by achievement interaction was significant.

Turning to the adjective list, separate analyses of variance were performed on the results summarized in Table 2 for trials to reach criterion and percentage of correct anticipations for Trials 2–12. For each measure, the only significant effect at beyond the .05 level is associated with achievement.

DISCUSSION

Ss high in EPPS n Achievement demonstrated significant superiority over low n Achievers in both verbal learning situations. This superiority of the high n Achievement group, particularly on the noun list which has not been associated with significant differences between high and low anxious (MAS) *Ss* in the past (Spence et al., 1956), would appear to be more consistent with an associative interpretation of n Achievement than a nonassociative one. Acceptance of a primarily associative interpretation would imply that the person high in n Achievement may be characterized by certain responses or habits that assist verbal learning. Assuming this interpretation to be appropriate, the present data do not provide a clear indication of what these responses might be. One may speculate, however, on the basis of an examination of the types of EPPS statements endorsed, that such responses as paying more attention, trying harder, etc. could characterize the person high in EPPS n Achievement. It is also possible, as Brown (1958) has suggested, that self-administered verbal directions, such as telling oneself to pay attention, to try to remember, etc., may contribute to response facilitation.

To check on the possibility that n Achievement may be associated with aptitude, ACE Psychological Composite percentile scores available for 13 *Ss* in the high n Achievement group and 16 *Ss* in the low n Achievement group were compared. Since sex did not appear to be a significant variable in the verbal learning results, sexes were combined. The mean ACE for the high n Achievement group was 69.46 (*SD* 18.59) and 58.70 (*SD* 22.81) for the low group. Analysis of variance revealed that the difference between groups was not significant (F = 1.77, *df* 1, 27).

Although the present findings are consistent with an associative conception, they do not permit, on the other hand, a clear rejection of the nonassociative view. As indicated earlier, the present adjective list has been used elsewhere with a presumed measure of drive, the Taylor MAS. On the assumption that the high association between each stimulus and response pair and the relative noncompetition among the pairs are associated with the correct responses being dominant during learning, high drive *Ss* have been expected to learn more rapidly than low drive *Ss*. Experimental findings have supported these expectations with this list (Spence et al., 1956). The present results, since they coincide with those obtained with anxious *Ss*, also lend themselves to a nonassociative interpretation of n Achievement.[3]

But, as suggested, the results with the noun list seem to temper a nonassociative view. It is possible, however, that n Achievement reflects both associative and nonassociative components, in that n Achievement scores may be associated with differential adaptation rates. Since all *Ss* in this experiment were exposed to the noun list first, it might be expected that high n achievers, if they are more adaptive, perhaps becoming more rapidly aware of what they are supposed to do, would learn more quickly than low n Achievers. The curves of learning for the high and low groups (Fig. 1A) on the noun list are consistent with this view, in showing the *Ss* high in EPPS n Achievement as initially (Trial 2) above the low group and maintaining this superiority over the five trials. Differential adaptation might be expected to affect performance on the noun list, but would be of considerably reduced significance on the subsequently administered adjective list. Thus, an associative conception would handle the differences on the noun list, while a nonassociative view would still be applicable to those obtained with the adjective list. On the whole, then, although the present findings indicate a relationship between EPPS n Achievement and verbal learning and are consistent with an associative interpretation, the presence of a nonassociative component as well cannot readily be discounted. Subsequent research might well consider either counter-balancing noun and adjective lists or employing both competitional and noncompetitional paired-associates learning materials to obtain a more complete picture of the theoretical properties of n Achievement.

REFERENCES

BROWN, J. S. *The motivation of behavior.* Unpublished manuscript, State Univer. of Iowa, 1958.

EDWARDS, A. S. *Manual: Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

---

[3] A sample of 130 college *Ss* yielded a product-moment correlation of .27, significant beyond the .01 level, between the Taylor MAS and EPPS n Achievement. The magnitude of this relationship does not provide very strong support for an energizing conception of achievement, since the MAS may well reflect other factors than drive. It may be noted, however, that this correlation differs considerably from the one of −.14 obtained by Edwards (1954) with 106 *Ss*.

FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

HEDLUND, J. L. Construction and evaluation of an objective test of achievement imagery. Unpublished doctoral dissertation, State Univer. of Iowa, 1953.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MCCLELLAND, D. C. Methods of measuring human motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society.* N. J.: Van Nostrand, 1958.

MELIKIAN, L. H. The relationship between Edwards' and McClelland's measures of achievement motivation. *J. consult. Psychol.*, 1958, **22**, 296–298.

SPENCE, K. W., TAYLOR, JANET A., & KETCHEL, RHODA. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received December 1, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

YEASAYERS AND NAYSAYERS:

AGREEING RESPONSE SET AS A PERSONALITY VARIABLE[1]

ARTHUR COUCH AND KENNETH KENISTON
Harvard University

PSYCHOLOGISTS have realized for a number of years that various response sets have considerable influence on most objective psychological test scores (Lorge, 1937). But only in the past few years, inspired by Cronbach's two articles (1946, 1950), has there been any systematic study of this biasing effect of response set and its implications for questionnaire validity.

Many of these investigations, recently summarized by Jackson and Messick (1958), have dealt with *agreeing* response set, defined as a tendency to agree (or disagree) with items regardless of their content. Psychological tests are frequently based on Likert-type scales with response alternatives ranging from Strongly Agree to Strongly Disagree. The agreeing tendency is of crucial importance with response scales of this type, since it is intrinsically involved in the method of measurement itself. In recent years, there has been a growing concern about the degree to which this agreeing set permeates the objective measures of many well known personality variables. Many studies have investigated the influence of agreement tendency on the "Authoritarian" F scale (Bass, 1955; Christie, Havel, & Seidenberg, 1958); on the variables on the MMPI (Fricke, 1956; Barnes, 1956), and on vocational interests (Roulon, 1931; Berdie, 1943; Holland, 1958).

A review of the literature reveals two general approaches to agreeing response set. From one viewpoint, response set is seen as primarily a statistical nuisance that must be controlled or suppressed by appropriate mathematical techniques. The agreement tendency is considered a rather superficial and ephemeral set, chiefly important as a mathematical factor that is compounded in scale measurements. While the statistical properties and consequences of the response set are given prominent attention, little interest is focused on the psychological determinants of the agreeing set in the individuals taking the tests. Cronbach, despite qualifications in other publications (1949, p. 51), summarizes this position in a recent discussion of accuracy of person perception: "In other instruments, the elevation score—the average response on all items—represents a trivial response set, such as the tendency to say 'yes' to both favorable and unfavorable questions. In that case, the dimension should probably be removed from the data" (1958, p. 366).

A second approach is represented by several researchers interested in agreeing response set as a manifestation of the responders' personality, interests, or personal "style" (Berdie, 1943; Jackson & Messick, 1958). Many of these investigators, though they view response set as a reflection of personality, postulate that agreement tendency is merely what it "appears" to be—a general "acquiescence" tendency that results in agreement with the rather general and authoritative statements of most questionnaires. Bass (1955) suggests that the response set is an indication of "social acquiescence"; while Leavitt, Hax, and Roche (1955) believe that the set is an "agreement with things authoritative." This view of response set as general acquiescence has led to the conclusion that many well-studied variables such as "authoritarianism" can be primarily explained in terms of the agreement phenomenon.

A somewhat different position is taken in the present study. Although response set is indeed a mathematical factor having major repercussions on objective test scores, exclusive concern with these statistical consequences may distract from the equally significant problem of the psychological determinants of response set. The basic hypothesis of this study is that response set is a manifestation of a deep-seated personality syndrome—whose underlying determinants serve to explain the

---

[1] This investigation was supported in part by research grant M1287 from the National Institute of Mental Health, USPHS. It was conducted at the Harvard Psychological Clinic Annex as part of a project directed by Henry A. Murray.

phenotypical phenomenon of "acquiescence" or "agreement."

This study attempts to analyze agreeing response set as a personality variable—using both objective test measurements and a program of clinical assessment. The first part of this report outlines the development of a reliable and valid measure of agreeing response set as it manifests itself in questionnaire items. We then show that this measure has the stability and generality which would characterize it as an index of an underlying personality dimension. This agreement measure was correlated against a series of objective test scales and items. The pattern of results suggested a set of hypotheses concerning the personality differences between individuals who tend to agree (yeasayers) and those who disagree (naysayers). The second part of this report presents a set of hypotheses about the personality dynamics behind the agreement tendency and attempts to validate these formulations by a clinical study of subjects with marked tendencies to "agree" and "disagree." Results of a structured interview and sentence-completion test given to these "yeasayers" and naysayers enabled us to verify and conceptualize more clearly the inner dynamics behind this "agreeing" personality syndrome. A general interpretation and discussion of the results obtained in both parts of the study is presented in the concluding section.

PART I: QUESTIONNAIRE STUDIES

The initial problem was to obtain a valid measure of response set, which we have defined as a general "tendency to agree or disagree with questionnaire items, regardless of their content." It is obviously important that the measure of response set not be "contaminated" by the undue presence of any specific cluster of content variables, as would certainly happen if the amount of agreement on any one or two standard psychological test scales were considered an adequate indicator of this general response tendency.

In order to create an appropriate measure of agreement tendency that was not affected by "content" variables, we decided that the questionnaire items included in it would have to fulfill three criteria: (*a*) the items should measure heterogeneous content variables from widely different psychological areas; (*b*) these content variables should be selected so that the total set is psychologically "balanced" to insure an over-all diffuseness of content; and (*c*) the total number of individual items in the over-all response measure should be very large (at least 300) so that the general response tendency has sufficient opportunity to manifest itself by permeating the answer pattern in a consistent way. With these purposes in mind, we began our study.

Development of the Over-all Agreement Score (OAS)

Subjects. As part of the assessment procedure in the selection of *S*s for a clinical research project, a large battery of objective personality tests was administered to 61 paid volunteer students from Harvard College. These *S*s had been selected on formal criteria (male, sophomore, satisfactory grades) from 200 volunteers enrolled in an undergraduate course in the social sciences.

Procedure. Most of the objective test items were answered in a series of three weekly sessions in the fall; a final group of tests was given later in the spring. The *S*s gave responses to 681 items, using a scale of the Likert type. The response categories were given values from 1 to 7, and were worded as follows:

(1)	*(2)*	*(3)*	*(4)*
Strongly Disagree	Disagree	Slightly Disagree	No Answer
(5)	*(6)*	*(7)*	
Slightly Agree	Agree	Strongly Agree	

The items included many from previously developed tests, such as Davids' Affect Questionnaire (1955), the factorial scales from Couch and Bales' Value Profile (1955), the Authoritarianism F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), Thurstone's Temperament Schedule (1949), Cattell's 16 P.F. Personality Inventory (Cattell, Saunders, & Stice, 1957), Rokeach's Dogmatism Scales (1956), and Bass' Social Acquiescence Scale (1956). In addition to these tests, well over 200 new items were included in an attempt to develop factor scales for a variety of psychological concepts in the areas of orality, anality, introspectiveness, psychological inertia, depression, elation, superego integration, insightfulness, self-concept, and attitudes towards American culture.

To control the effect of content variables in our measure of agreeing response set, we based it entirely on 360 selected items out of the total of 681 items of the Likert-scale type. In an attempt to fulfill our criteria for a response measure unaffected by specific content, we chose items from a large number (30) of heterogeneous scales that had "psychological" opposites, such as Trust-Distrust, Dependency-Self-sufficiency, Rationalism-Religiousness, Optimism-Pessimism, etc.

By including an equal number of items on "opposite" ends of each of the dimensions measured, we hoped to "cancel out" the effect of any content variable on the Over-all Agreement Score. This diversity and "psychological balancing" of content reduced the possibility that the measure of agreeing tendency would be unduly influenced by some extraneous pattern of personality variables—a total score over all items makes no psychological sense. To the extent that this "content cancelling" has been achieved, the sum score of all 360 item responses is a pure measure of agreeing response set, that is, of the *S*'s tendency to agree (or disagree) with items *regardless* of content. This approach should also reduce the possibility that correlations between the agreement measure and other scales or items are produced by content variables unwittingly built into the measure itself.

An Over-all Agreement Score (OAS) was computed for each *S* by taking the mean of their responses to the 360 items. The range of scores was from 3.1 to 4.5, with a mean value of 3.9 and a standard deviation of 0.3. The first sign of the importance of this measure was indicated by the high (+.85) Spearman-Brown split-half (even-odd) reliability of the entire 360 item scale. The OAS thus provides a reliable measure of agreeing response set as defined in this study.

Generality of the OAS

Having established the statistical reliability of the OAS, it is necessary to demonstrate that this measure has empirical characteristics which support our contention that it measures a generalized agreement tendency. Since the OAS was based on an arbitrary selection of items, we considered that it was necessary to test its generality by correlating it with the grand sum of all 681 items of the Likert type in our questionnaire battery. This "grand sum" correlated +.94 with the OAS, a high value even considering the 53% overlap of items. Although the larger number of items also appears to serve the function of guaranteeing heterogeneity and content "elimination," in this study we have used the OAS as the scale that most fully meets our criteria for a valid measure of response set.

To test the extent that response set varied from one testing session to the next, we computed the correlations between the sum of "content-cancelled" items (120) from the first two weeks of testing and the sum of "content-cancelled" items (240) from the third week of testing. The correlation coefficient was +.73, indicating that agreeing and disagreeing response set operates fairly consistently over time, and is not merely the product of the *S*'s momentary mood.

TABLE 1
CORRELATIONS OF RESPONSE CATEGORIES WITH OAS

Likert Response Categories	Percentage Frequency (Mean)	Percentage Frequency (*SD*)	Correlations with OAS ($N = 61$)
Strongly Agree	4.8	3.3	+.41
Agree	17.1	4.6	+.55
Slightly Agree	25.5	5.4	+.23
No Answer	0.2	0.6	−.10
Slightly Disagree	22.8	6.0	−.23
Disagree	21.8	4.9	−.54
Strongly Disagree	7.9	4.0	−.38

The OAS as a mean score indicates the average point along the Likert scale around which the *S* tends to center his responses. However, this mean score does not tell us what actual response category preferences are associated with high and low OAS scores. To investigate these relationships, we computed the percentage frequency with which the *S*s choose each of the six Likert response categories and then correlated these figures with the OAS scores. The mean and standard deviation of these percentage frequencies are presented along with the correlation coefficients in Table 1.

These results demonstrate that the OAS is most highly related to a response preference for the categories of Agree (in a positive way) and Disagree (in a negative way). These correlations are affected by the relative contribution to the OAS of each response category—as reflected in the Mean Percentage Frequency. This "part-whole" effect may explain the lower correlations for Strongly Agree and Strongly Disagree, but since other response categories have about the same mean frequency, the uniquely high correlation of the Agree and Disagree categories point to a real relationship of response preference and the agreement tendency. The preference for definite "agreement" responses supports our interpretation of the OAS measure. It should also be noted that the frequency of using *disagreement* categories is an equally prominent indicator of the agreement tendency—only the direction of scoring is reversed. Thus, to label the response tendency *agreement* is only an arbitrary preference on our part; the opposite *disagreeing response tendency* would be an equally appropriate conceptualization.

Yeasaying and true-saying. If this tendency to agree or disagree is a manifestation of a general response set, this set should also permeate objective tests that do not employ Likert-scale items but use response categories that are similar in psychological meaning to agreement and disagreement. In order to test the generality of response set, we investigated the relationship of the OAS with the agreement equivalent (marking true) on the MMPI (Minnesota Multiphasic Personality Inventory). About six months after the initial testing sessions, the *S*s filled out the entire 566-item MMPI. As a measure of a "true-saying" response tendency, we computed a Sum True score for each *S* by counting the total number of True responses on all items. This score, though perhaps also reflecting dispositions other than response set, had a correlation of +.64 with the OAS. A correlation of this magnitude indicates that response set influences answers both to Agree-Disagree Likert items and to True-False answer patterns.

Yeasaying and yes-saying. To test further the generality of agreeing response set in the objective test domain, we examined the characteristic response pattern of our *S*s on another standard personality inventory—Cattell's 16 P.F. Test. This test consists of Forms A and B, with a grand total of 374 items, of which 217 (about 60%) have an answer pattern which offers the choices: Yes, In Between, No. We counted the total number of times each *S* had marked Yes on those 217 items where that response was possible. This Sum Yes score was then correlated with the OAS, giving a correlation of +.56. This finding adds support to our hypothesis that agreeing response set is a disposition of considerable influence and stability, and indicates again that the OAS functions as a reliable and valid measure of this general agreeing tendency.

Correlates of the OAS

The results obtained in developing the Overall Agreement Score, namely the high reliability, stability over time, and generality of agreeing response set, support the view that this set is the manifestation of a personality syndrome. We therefore proceeded to treat the OAS as a criterion measure of this agreeing response set, and correlated the score with a series of objective test domains. This step-wise search for the correlates of the agreement response set gradually built up a consistent and meaningful pattern which characterized the personality of individuals who tend to "agree" (or conversely "disagree") on questionnaires regardless of item content.

Relationship to the MMPI. There are intrinsic difficulties in the task of relating a measure of agreeing response tendency to any series of objectively measured personality variables; for the response tendency itself is an *integral part of the measurement technique* involved in assessing the personality variable. This is especially true when the variable measured uses items scored mainly in one direction, when, for example, some variable X is scored by counting all items marked Agree, True, or Yes.

On the MMPI we have already noted that the tendency to mark True is highly related to the general agreeing response set. For each of the 32 MMPI scales[2] that were computed for each *S*, a tabulation was made of the percentage of items in the scale which count towards a high score on that scale when marked True. All scales were then correlated with the OAS. These correlations and the Percentage True counts are presented in Table 2.

The analysis presented in Table 2 illustrates the very close relationship between the amount of agreement response component inherent in the scale as measured by the Percentage True count and the correlation of that scale with our measure of agreement tendency. Scales that rely primarily on True responses for a high score tend to have positive correlations with the OAS, while scales relying on False responses for a high score tend to have negative correlations. In fact, over the 32 standard scales, these two measurements have a rank order correlation of +.86, indicating that scales with an unusually high (or low) percentage of items scored True are significantly influenced by the agreeing response set.

Because of this possible source of "contamination," an interpretation of the psychological implications of the correlations between the MMPI scales and the OAS measure requires caution. The pattern of scales having significant positive and negative correlations with the OAS clearly suggests pervasive psy-

---

[2] Fuller descriptions of each scale and appropriate reference are available in the article by Kassebaum, Couch, and Slater (1959).

TABLE 2

CORRELATIONS OF THE MMPI SCALES WITH THE OVER-ALL AGREEMENT SCORE (OAS)

Scales		Percentage "True"	Correlation with OAS ($N = 61$)
ST	Sum True	100	+.64**
Pr	Prejudice	88	+.51**
A	Anxiety	97	+.50**
Dp	Dependency	86	+.48**
X2	Impulsivity	90	+.46**
Pt	Psychasthenia	81	+.45**
Ma	Mania	76	+.41**
Sc	Schizophrenia	74	+.38**
Fm	Feminine Masochism	66	+.33*
Hs	Hypochondriasis	33	+.29*
D	Depression	35	+.22
F	Validity Scale	69	+.20
Do	Dominance	25	+.15
Pd	Psychopathic	48	+.15
Si	Social Introversion	49	+.07
Pa	Paranoia	63	+.06
Sp	Social Ability	68	+.03
Mf	Feminine Interests	47	+.01
X1	Social Presence	25	−.06
St	Social Status	44	−.07
Hy	Hysteria	22	−.08
Oi	Originality	16	−.26*
Ie	Intellectual Efficiency	38	−.28*
Lp	Leadership	28	−.31*
R	Repression Denial	0	−.34**
Es	Ego Strength	37	−.35**
Ac	Achievement via Conformance	22	−.35**
K	Suppressor Scale	3	−.38**
Rp	Role-playing Ability	41	−.40**
L	Lie Scale	0	−.42**
Ai	Achievement via Independence	19	−.44**
Re	Social Responsibility	28	−.48**
To	Tolerance	10	−.52**

* .05 level of significance ($r = .26$).
** .01 level of significance ($r = .33$).

chological differences between yeasayers and naysayers. However, in so far as these correlations are related to the Percentage True counts of the scales, the importance of the interpreted "content" of these MMPI variables may seem to be decreased. With the exception of the few scales with approximately 50% True counts—such as (*Ie*) Intellectual Efficiency, (*Es*) Ego-strength, and (*Rp*) Role-Playing Ability—none of the variables with significant correlations are free of the "contaminating" effect of response set on their level of relationship with the OAS—thus presenting difficulties in using the correlations to identify the "psychological" characteristics of the agreement response tendency.

Nonetheless, it should be remembered that most MMPI scales were developed by the use of "criterion" groups, and therefore the "psychological" nature of agreement response set would have contributed to the empirical pattern of item responses that differentiated individuals *high* vs. *low* on the variable of the scale. We can probably infer that the original populations of *S*s high and low on Prejudice, Psychasthenia, Mania, Achievement via Independence, and the other concepts of the empirical scales had personality differences related to an "agreement" or "disagreement" response tendency. In other words, the Percentage True of the empirical scale reflects personality differences in the criterion groups that are related to the response tendency. To the extent that this inference is valid, the correlational results can be helpful in identifying the personality correlates of response set, despite the apparent "contamination" of the scale scores with response set.

Table 2 shows a number of significant correlations of MMPI scales with the OAS. The positive correlations of (*X2*) Impulsivity, (*Dp*) Dependency, and (*Ma*) Mania with the OAS suggest that individuals high on the agreeing tendency show signs of an extrovertive impulsivity. The centrality of this general dimension seems borne out by the cluster of high negative correlations with the variables: (*R*) Repression-suppression, (*Re*) Responsibility, (*Ai*) Achievement via Independence, and (*Ac*) Academic Achievement via Conformance. All these scales point to introvertive intellectual interests, inhibitions, and a control of impulsivity. Another dimension which seems indicated is that of ego strength. Yeasayers are high on such scales as (*A*) Anxiety and (*Pt*) Psychasthenia, while naysayers obtain high scores on the scales of (*Es*) Ego Strength and the (*K*) Supressor scale. In terms of a recent factor analytic study of the MMPI (Kassebaum, Couch, & Slater, 1959) this pattern of results places the agreeing response tendency along the axis of a fusion factor conceived as Impulsivity vs. Intellectual Control. The fact that the (*Pr*) Prejudice scale has the highest correlation (+.51) with the OAS and (*To*) Tolerance scale the lowest correlation (−.52) may be primarily explained by this main factorial dimension. Further implications of these findings are best discussed after more of the correlates of the OAS are considered.

TABLE 3

CORRELATIONS OF THE CATTELL 16 P.F. SCALES WITH THE OAS SCORE

Scale	Correlation with OAS ($N = 61$)
A —Warm, Sociable vs. Aloof, Cold	+.27*
B —Intelligent vs. Dull	+.07
C —Mature, Calm vs. Emotional, Unstable	−.35**
E —Dominant, Aggressive vs. Submissive	+.22
F —Enthusiastic, Talkative vs. Silent	+.23
G —Conscientious vs. Undependable	+.15
H —Adventurous vs. Timid	+.02
I —Sensitive vs. Tough	+.02
L —Suspecting, Jealous vs. Trustful	+.39**
M —Eccentric vs. Conventional	+.10
N —Sophisticated vs. Simple	+.19
O —Insecure, Anxious vs. Unshakable	+.38**
Q1—Critical vs. Conservative, Accepting	−.02
Q2—Self-sufficient vs. Dependent	−.17
Q3—Controlled, Exact vs. Lax, Unsure	−.35**
Q4—Tense, Excitable vs. Phlegmatic, Poised	+.38**

* Significant at the .05 level.

** Significant at the .01 level.

TABLE 4

CORRELATIONS OF THURSTONE'S TEMPERAMENT SCALES WITH THE OAS SCORE

Scale (Factor)	Percentage Scored "Yes"	Correlation with OAS ($N = 61$)
	%	
A—Active	65	+.13
V—Vigorous	100	+.01
I —Impulsive	90	+.19
D—Dominant	70	+.14
E—Stable	35	−.38*
S—Sociable	80	+.10
R—Reflective	95	+.16

* Significant at the .01 level.

Relationship to the Cattell and Thurstone Personality Tests. In Tables 3 and 4, we have presented the correlations of the OAS with each of the 16 standard scales (Forms A and B combined) in the Cattell 16 P.F. Test and with the seven factor scales of Thurstone's Temperament Schedule. The percentage of Yes responses is presented for the Thurstone scales. These figures are not appropriate for the Cattell test which uses a large number of (a) vs. (b) alternative choices as item responses.

In this battery of tests, the highest positive correlations with the OAS are with Cattell's scales (L) Suspecting, Jealous vs. Trustful, (O) Insecure, Anxious vs. Unshakable, and (Q4) Tense, Excitable vs. Phlegmatic, Poised. The highest negative relationships are with Cattell's (C) Mature, Calm vs. Emotional, Unstable, (Q3) Controlled, Exact vs. Lax, Unsure. In the Thurstone test, there is a significant negative correlation with the (E) Stable temperament scale—which also has the lowest percentage of Yes responses involved in its score. We note again the theme of emotional control vs. impulsivity. The naysayers present a picture of calm, controlled stability; while yeasayers seem more emotionally determined, openly admitting anxiety and distrustfulness. The highest correlation (+.39) is with the Suspecting, Jealous vs. Trustful Scale, indicating that a *lack* of trust is associated with the OAS. This datum reminds us that one should not too quickly generalize from the apparent "acquiescence" of the yeasayer.

It is interesting to note the lack of relationship of the OAS to intelligence as measured by the (B) Intelligent vs. Dull scale (+.07). This pattern was verified by nonsignificant correlations of the OAS with SAT (Verbal Aptitude Subtest) (−.04) and MAT (Mathematical Aptitude Subtest) (.00) of the College Entrance Examination Board scores. In our population, it appears that intelligence is not an important aspect of agreeing response set.

Relationship to Basic Disposition Scales. As part of other research at the Harvard Psychological Clinic Annex, we had developed a set of factorial scales to measure basic dispositions in the areas of Orality, Anality, Alienation, and Psychological Inertia. In a series of unreported studies large domains of test items in each of these four areas were factor analyzed. Several reliable factor scales were developed, each possessing the statistical characteristics of factorial purity and domain orthogonality. For some variables, both Positive and Balanced scale measures are available. These latter scales are "balanced" in the sense that an equal number of items are scored "positively" and "negatively." Thus, Balanced scales should serve to cancel out any contamination due to agreeing response set. The Positive factor scales, reported below, consist of items scored only in one direction. Table 5 presents the correlations of the OAS with these Basic Disposition Factor Scales, subdivided into Balanced and Positive scales.

This table permits us to compare the correlations of Balanced and Positive scales designed to measure the same concept. In every

TABLE 5

CORRELATIONS OF THE BASIC DISPOSITION FACTOR SCALES WITH THE OAS

Basic Disposition Factors	Correlations with OAS ($N = 61$)
Positive Scales:	
Conventional Anxiety	+.57**
Conscious Hostility	+.46**
Alienation	+.44**
Distrust	+.32*
Pessimism	+.18
Optimism	+.17
Trust	−.07
Oral Aggression	+.43**
Oral Dependency	+.40**
Oral Involvement	+.14
Oral Rejectiveness	+.03
Oral Counterdependency	−.03
Anality Total Scale	+.75**
Anal Preoccupation	+.75**
Anal Submission	+.67**
Anal Rebellion	+.61**
Anal Concern with Power	+.55**
Anal Orderliness	+.25
Anal Messiness	+.20
Secondary Process Inertia—Excitability	+.57**
Low Emotional Inertia	+.56**
Equilibrium Inertia—Emotional Variability	+.48**
Inconstancy	+.37**
Quick Reactions	+.33*
Perseveration	+.17
Slow Reactions	−.02
High Psychological Inertia	−.20
Balanced Scales:	
Distrust vs. Trust	+.20
Optimism vs. Pessimism	+.01
Oral Dependency vs. Counterdependency	+.25
Anal Orderliness vs. Messiness	+.07
Emotional Inertia: High vs. Low	−.43**
Equilibrium Inertia: Emotional Stability vs. Variability	−.33*
Secondary Process Inertia—Impulse Control vs. Excitability	−.30*
Quick Reactions vs. Slow Reactions	+.20
Inconstancy vs. Perseveration	+.10

* Significant at the .05 level.
** Significant at the .01 level.

case where the comparison is possible, the correlations of Balanced scales with the OAS are *less* positive than that of their Positive scale counterparts, the average drop being .18. This difference is of course statistically determined by the fact that the agreeing response set is actually built into all Positive scales and systematically cancelled from Balanced scales. The comparison again points out the important contribution that agreeing response tendency makes to the variance of Positive scales, and demonstrates that "balancing-out" response set reduces the response set component in the scale score.

With this effect of response set on the variables kept in mind, the pattern of correlational results in Table 5 brings out new facets of the agreeing response tendency. Several of the recently developed Basic Disposition Factor scales have significantly positive correlations with the OAS. Many of these scales were derived from our factor analysis of the Anality domain. Since these scales are not yet described in the literature, we have included here a characteristic item for the scales mentioned. It can be seen that the highest positive correlations are for the dispositional scales of Anal Preoccupation ("I enjoy exploring slums and alleys, dirty though they are"); Anal Rebellion ("I could really shock people if I said all of the dirty things I think"); and Anal Submissiveness ("People need someone strong to lean on"). These results suggest that the previously noted relationships of the OAS to impulsivity and rejection of emotional control may be part of a more basic dispositional complex centered around anality. We also note a positive relationship to the scales measuring Alienation ("The idea of adjusting to society as now constituted fills me with horror"); and Conventional Anxiety ("There are few more miserable experiences than going to bed night after night knowing you are so upset that worry will not let you sleep"). These variables seem related to the general maladjustment and distrustfulness noted in the MMPI scales and in the Cattell factors.

On the opposite end of the agreeing tendency, a group of Balanced scales in the area of *psychological inertia* is found to characterize the naysayers. The factor scales with the highest negative correlations with the OAS are: Secondary Process Inertia ("I have to think things through thoroughly before I act"); Equilibrium Inertia ("Uncontrolled impulsiveness is not part of my make-up"); and Emotional Inertia ("Whatever emotional mood I have is likely to stay with me for a long time"). This pattern of general retention of impulses and the reluctance to respond quickly to situational demands may very well represent the opposing reaction to the anal expressive syndrome pointed out by the cluster of positively correlated scales, and contribute a new meaning to the general emotional stability indicated by the previously reported correlational results. It may be that standard measures of ego strength, adjustment, and maturity, when used on populations of normals, are related primarily to this kind of pervasive control of impulses by dominant secondary processes.

Relationship to Authoritarianism. Much of the previous research on agreeing response set has focused on its relationship with authoritarianism. It has been argued that scores on the F scale are in large part a function of this agreeing tendency, with the implication that the original formulations concerning the authoritarian personality may be invalidated by this fact. To investigate this question, we made an especially intensive study of the relationship between the F scale and the OAS.

The literature on authoritarianism and "acquiescence" often concentrates on the question of what percentage of F scale variance can be attributed to the agreeing response tendency. With the OAS, a fairly direct measure of this percentage can be obtained from the correlation of the OAS with the F scale. But many objective measures of authoritarianism suffer from a lack of knowledge of their dimensional purity. In an effort to avoid these difficulties, we used as measures of authoritarianism two scales that had been developed by Couch and Bales (1955) in their factor analysis of the Value Profile. These scales are factorially pure measures of authoritarianism, both having very high saturation on the first factor of their analysis which uniquely defines a 30-item F score with a loading of +.90. One scale, called the Positive F scale, consists of 10 items all scored in the positive direction: for example, "Obedience and respect for authority are the most important virtues children should learn." The other scale, the Balanced F scale, consists of 5 items scored positively and 5 items scored negatively. The factorial purity of this scale is achieved by a balancing out of other factor components. Thus, negative items that are not pure "reversed" authoritarianism can be used when properly compensated for by complementary factorially "impure" positive items. With this Balanced F scale, agreeing response set is effectively cancelled out; thus, the correlation with the OAS should reflect the lack of this "contaminating" influence. The results of correlating these two factorially "equal" measures of authoritarianism with the OAS are presented in Table 6.

TABLE 6

CORRELATIONS OF AUTHORITARIANISM MEASURES WITH THE OAS

Factor Scales	Correlations with OAS ($N = 61$)
Positive F scale	+.37*
Balanced F scale	+.09**

* Significant at the .01 level.

** Significantly different from the above $r = +.37$ at the .01 level.

As Table 6 indicates, the Positive F scale correlates significantly with the OAS, supporting the view that the F scale, if measured by positive items, is influenced to some extent by agreeing response set. But the percentage of variance attributed to this agreement effect is only 14%, if we assume that the correlation with the OAS is an appropriate measure of this influence. More substantial evidence that "pure" authoritarianism is actually uneffected by agreement tendency comes from the nearly zero (+.09) correlation between the Balanced F scale and the OAS. Since both F scales are highly loaded on the authoritarianism factor and are significantly intercorrelated (+.56), these results suggest that the agreeing response tendency is not related to "pure" authoritarianism when its measurement is not inherently contaminated by response set. In another section, we return to this issue in a more substantive context where a factor analysis supports the same conclusion.

Relationship to Social Acquiescence and Dogmatism. We turn to the areas of social acquiescence and dogmatism where previous research had suggested certain relationships to response set. Bass (1956) developed a Social Acquiescence Scale in a manner somewhat similar to the method we used in obtaining our Over-all Agreement Score, although only aphorisms were involved in his scale. One-half of Bass' scale (28 odd-numbered items) was included in our original test battery. In addition to these items, we gave Rokeach's (1956) Dogmatism Scales, which consist of 40 items measuring different aspects of ideological and functional dogmatism. Table 7 presents the correlations of these scales with the OAS.

TABLE 7

CORRELATIONS OF SOCIAL ACQUIESCENCE AND DOGMATISM WITH THE OAS

Scales	Correlations with OAS ($N = 61$)
Bass' Social Acquiescence Scale	+.30*
Rokeach's Dogmatism Scales	
Structure scale	+.26*
Content scale	+.29*
Function scale	+.42**
Total scale	+.40**

* Significant at the .05 level.

** Significant at the .01 level.

The Social Acquiescence Scale correlates significantly (+.30) with the OAS, but at a lower level than would be expected if "acquiescence" were actually a central variable in the agreeing response tendency. This rather low correlation is especially indicative, since all items are scored in the direction of agreement. Given the general pattern of positive correlations between other Positive scales and the OAS, this result suggests that the Social Acquiescence Scale is not distinctively related to agreeing response set.

With the Dogmatism Scales, we found the highest correlation between the Function Scale and the OAS (+.42). The items here are of the type: "I would like it if I could find someone who would tell me how to solve my personal problems." This scale attempts to measure the underlying function of dogmatism in the individual's personality, especially in controlling the fear of aloneness and apprehension about the future. The relationship between the Dogmatism Scale and the OAS appears to reflect the diffuse anxiety of the yeasayer, and to suggest that he seeks dependency on external figures for support.

Development of the Agreement Response Scale

As the next stage of our research we correlated the OAS to each of the 681 items in our questionnaire battery. The item-to-OAS correlations ranged from −.21 to +.54; the mean correlation for all items was +.17. The high mean correlation demonstrates again the extent to which response set has influenced all item responses. This trend was also apparent from the fact that 88% of all items correlated positively with the OAS. Well over 100 items had correlations with the OAS greater than +.30. Table 8 lists a sample of items with the highest positive relationship to the response tendency, along with the item means, standard deviations, and correlations with the OAS.

These items with the highest positive correlations serve to identify the agreement response tendency on the most specific level of separate statements. Since the average of

TABLE 8

CORRELATIONS OF INDIVIDUAL ITEMS WITH THE OAS SCORE

Correlation with OAS ($N = 61$)	Mean	*SD*	Items[a]
+.54	4.9	1.4	Novelty has a great appeal to me.
.54	4.5	1.7	I crave excitement.
.51	3.5	1.6	It's a wonderful feeling to sit surrounded by your possessions.
.48	3.3	1.6	There are few things more satisfying than really to splurge on something—books, clothes, furniture, etc.
.47	4.2	1.7	Only the desire to achieve great things will bring a man's mind into full activity.
.46	3.6	1.6	Nothing is worse than an offensive odor.
.45	3.6	1.3	In most conversations, I tend to bounce from topic to topic.
.44	2.9	1.6	I really envy the man who can walk up to anybody and tell him off to his face.
.44	4.1	1.8	I could really shock people if I said all of the dirty things I think.
.42	4.3	1.7	There are few more miserable experiences than going to bed night after night knowing you are so upset that worry will not let you sleep.
.42	3.8	1.6	I tend to make decisions on the spur of the moment.
.41	3.8	1.8	Little things upset me.
.41	3.0	1.3	Drop reminders of yourself wherever you go and your life's trail will be well remembered.
.40	2.1	1.3	I like nothing better than having breakfast in bed.
.40	4.2	1.6	My mood is easily influenced by the people around me.

[a] The 15 items listed here we now consider the best short scale measure of the agreeing response tendency.

these 15 item means (3.7) and of item standard deviations (1.6) is almost identical to the average mean for all 681 items (3.8) and their average standard deviation (1.5), no purely statistical factors are determining the correlational results. The items reflect to a great extent the same array of variables that we have already noted—impulsivity, anality, and instability. A more detailed analysis of these implications is presented in the final stages of the scale development.

A selection of 20 items with the highest correlations to the OAS constituted our first approximation to an Agreement Response Scale (ARS). The split-half reliability of this scale was very satisfactory (+.86). With all items being scored in a positive direction, the response set is, of course, working in our favor—as is uniquely desired in this case. This 20-item scale (ARS) correlates +.79 with the OAS, thus giving it considerable criterion validity.

However, it is obvious that all of the statistical uniqueness of our single sample is contributing to this pattern of results for the ARS. We felt it necessary to test the internal consistency of the tentative scale by administering it to an independent population. A 120-item questionnaire was developed which included this ARS scale, as well as an array of items that we hoped would bring out new facets of the agreeing tendency. We also included a number of standard F scale items and several new ones written to measure "low" authoritarian attitudes.

The questionnaire was given to 54 students in an upper-level psychology course. For this new sample, the split-half reliability of the ARS remained quite high (+.72), indicating that our measure had the characteristics of internal consistency. As a next step in scale development, we did a factor analysis of the 120 items. These items were intercorrelated and the resulting matrix factor analyzed by the Thurstone complete centroid method. Four factors were extracted, which altogether accounted for 28% of the item domain variance. By Thomson's (1939) Pooling Square technique, we computed the factor loadings of the total 20-item Agreement Response Scale, and then made orthogonal rotations so as to define the ARS by the first factor.

TABLE 9

FACTOR LOADINGS OF ITEMS ON AGREEMENT RESPONSE SCALE DIMENSION

Factor I Loading	Item
+.66	Let us eat, drink, and be merry, for tomorrow we die.
+.63	I tend to make decisions on the spur of the moment.
+.61	There are few things more satisfying than really to splurge on something—books, clothes, furniture, etc.
+.57	Here today, gone tomorrow—that's my motto!
+.50	Novelty has a great appeal to me.
+.49	My feelings about others fluctuate a good deal.
+.48	Conscience is another name for fear.
+.46	Movement, travel, change, excitement—that's the life for me!
+.45	I'm apt to really blow up, but it doesn't last long.
+.42	It's great fun just to mess around.
+.41	There's nothing so satisfying as really to tell someone off.
+.40	I really enjoy plenty of excitement.
−.61	One should not give free rein to the passions, but rather control and weigh them before expressing them.
−.51	I seldom if ever lose my temper.
−.49	Uncontrolled impulsiveness is not part of my make-up.
−.37	It's hard to get me upset.
−.34	My speech is quite slow and deliberate.
−.33	I feel uncomfortable when people get too emotional.
−.33	I almost never respond impulsively to people or events.
+.86	Agreement Response Scale

With the assumption that the ARS remained a valid measure of agreeing response set, the first factor of this new analysis may be considered a dimensional representation of our central variable. In Table 9 we have listed those items with the highest positive and negative loadings on Factor I, which might best be labelled the "Agreement Response Scale" dimension.

A difference in emotional responsiveness seems to be the central concept of the agreement dimension. Items with high positive loadings on this Agreement Response Scale factor tend to emphasize an impulsive acceptance of immediate internal stimuli ("I'm apt to really blow up, but it doesn't last long") and the desire for external stimuli in the form of emotional sensations and thrill-seeking ("Novelty has a great appeal to me"). This pattern might be summarized by the term "stimulus acceptance," indicating a tendency to enjoy and yield to immediate, situational, short-term forces of both an internal and external nature. The items with high negative loadings stress just the opposite, namely, rejection of these same emotional and environmental pressures. The central theme of the naysaying position is the effort to control and minimize internal and external forces seeking expression. Naysayers emphasize ego control of libidinal impulses ("One should not give free rein to the passions, but rather control and weigh them before expressing them") as well as delay and inhibition of responses to provocative press in the environment ("I am not easily provoked").

Importance of Item Tone. An examination of items with high loadings on the agreement factor suggested the importance of item tone in the distinction between yeasaying and naysaying characteristics. The positive items indicated that the yeasayers' impulsiveness was revealed most clearly in statements whose *item tone* was extreme and enthusiastic. Exclamatory phrases and colloquial slang seemed to have intrinsic appeal to the personality of the yeasayer: for example, "Here today, gone tomorrow—that's my motto!" and "It's great fun just to mess around!" On the opposite end of the agreeing response dimension, we noted that items with negative loadings were phrased in guarded, qualified, and cautious language. Items such as "I seldom if ever lose my temper," and "I almost never respond impulsively to people or events," point out the degree that "careful," "realistic," and intellectually qualified phrasing are inherent in the item tone of naysaying items. We feel that this aspect of item phraseology is highly related to the agreement response tendency on questionnaires. If the item has an enthusiastic, colloquial tone, it will serve to attract the yeasayer and to repel the naysayer. On the other hand, a qualified and hesitant statement will appear reasonable and bring forth agreement from the generally disagreeing naysayer, while provoking the generally effusive yeasayer to one of his infrequent disagreements. Although the psychological content of "expression" vs. "control" of emotional impulses certainly influences the language of the statements used for measurement, the characteristic phraseology of items high on this factor suggests that *item tone* itself contributes to eliciting special agreement from yeasayers and disagreement from naysayers.

The factor analysis serves to support the conclusions we had arrived at from the previous investigation of the empirical correlates of the Over-all Agreement Score in the realm of objective personality scales. The factor dimension of the Agreement Response Scale shows a conceptual similarity to the pattern of variables that were highly correlated to the OAS.

Factorial Relationship of Agreement Response Scale to Authoritarianism. A number of items in the area of authoritarianism were included in this factorial study in order to clarify further the relationship of agreeing response tendency to this variable. In addition to the regular "positive" F scale items, we added several new items designed to measure "low" authoritarianism in an agreement direction. The second factor was rotated through the F scale measures. This two-dimensional factor pattern of Agreement and Authoritarianism is presented in Fig. 1.

The factor plot of item loadings on Factors I and II shows that the dimension of the Agreement Response Scale was orthogonal to the dimension of authoritarianism. The Positive F scale items were scattered around the plus end of Factor II, while our new Negative F scale items defined the opposite pole. Several items were loaded on both main factors and these served to define the "fusion" locations of the factor space. The ARS items were split into two types by the authoritarianism dimension; those expressing a Desire for External Stimulation were associated with high authoritarianism, while items favoring an Expression of Unacceptable Internal Impulses were associated with low authoritarianism. On the opposite end of Factor I, items measuring the Disagreement tendency were also divided into two different types. High authoritarianism was reflected in items favoring a Control of Unacceptable Internal Impulses; while low authoritarianism was reflected in items favoring an Avoidance of External Stimulation.

The factor pattern of loadings illustrated in Fig. 1 supports our earlier finding that "pure" authoritarianism is independent of the agreeing response set. The measure of the agreement tendency, the Agreement Response

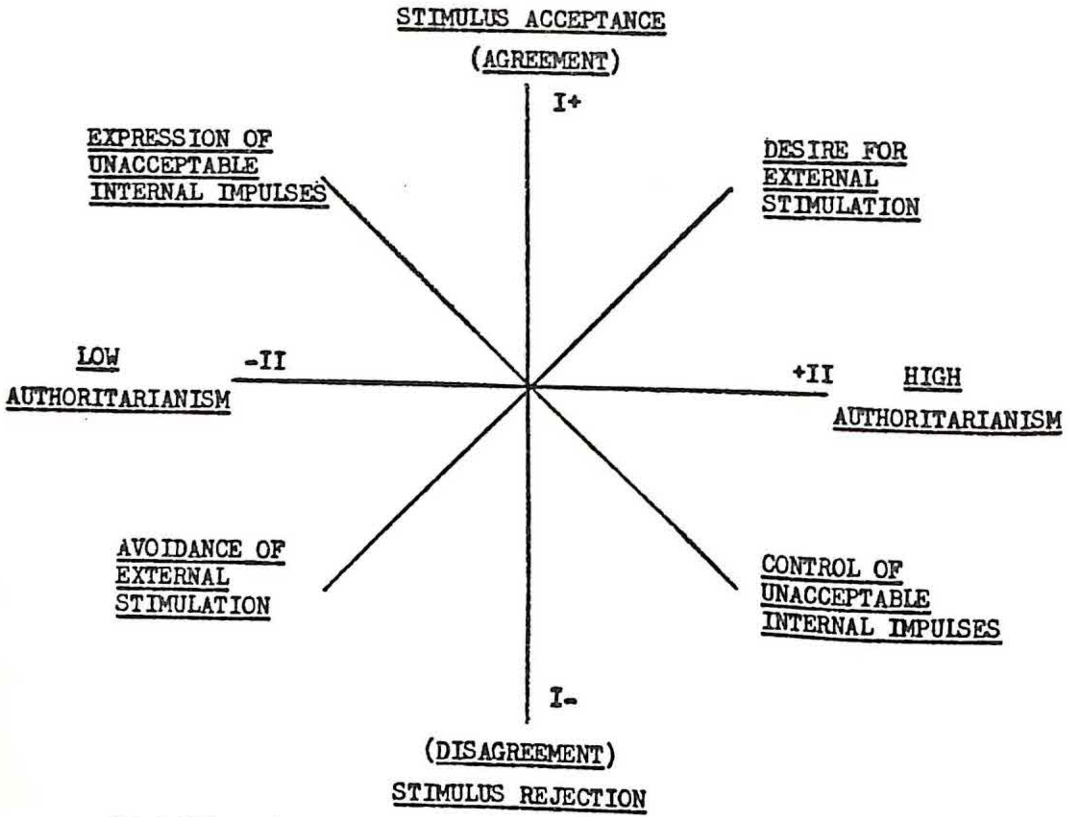

STIMULUS ACCEPTANCE
(AGREEMENT)
I+
EXPRESSION OF UNACCEPTABLE INTERNAL IMPULSES
DESIRE FOR EXTERNAL STIMULATION
LOW AUTHORITARIANISM
-II
+II
HIGH AUTHORITARIANISM
AVOIDANCE OF EXTERNAL STIMULATION
CONTROL OF UNACCEPTABLE INTERNAL IMPULSES
I-
(DISAGREEMENT)
STIMULUS REJECTION

FIG. 1. Schematic representation of "agreement" and "authoritarianism" factors.

Scale, uniquely defines the first dimension that is orthogonal to the second dimension of "authoritarianism." Although some of the agreement items had a component of *desiring stimulation*, while others reflected *expressing unacceptable impulses*, these items taken together balanced out the second factor and left the ARS a "pure" measure of response set. The very fact that the ARS contains these two complementary external and internal components supports our previous interpretation that Stimulus Acceptance vs. Stimulus Rejection represents the more abstract central concept involved in agreement response set.

Development of "Low" F Scale. Our research on the agreement response tendency suggested a new approach for the construction of a "reversed" or "low" F scale. It is clear that if items on opposite ends of the authoritarian dimension all contain a substantial component of the response set variable, then these "opposing" scales will tend toward a positive relationship regardless of the reversal of authoritarian content which would tend to cause a negative relationship. Other researchers (Bass, 1955; Jackson & Messick, 1958) have tried this type of "content reversal" with only limited success in obtaining negatively related scales. Our research has pointed out the importance of "item tone" in the agreement response tendency. Since the phraseology of an item contributes to the amount of agreement or disagreement which it attracts, we hypothesized that by manipulating the item tone of low authoritarian items so as to attract naysayers and repel yeasayers, we could increase the negative relationship between the "high" and "low" authoritarian scales. Several "nonauthoritarian" statements were therefore phrased in a "naysaying manner" that emphasized qualification and a lack of exclamatory enthusiasm. The five items of this kind with the highest negative loadings in the previous factor analysis formed a new "Low F scale." These "naysaying" type items are listed below.

"Naysaying" Low F scale

1. People tend to place too much emphasis on respect for authority.
2. It may well be that children who talk back to

their parents actually respect them more in the long run.

3. I seldom have any enthusiasm for respect and obedience for authority.

4. It would be preferable if there were less enthusiasm about strict discipline, rugged determination, and the will to work and fight for family and country.

5. I would not myself consider patriotism and loyalty to be the first requirements of a good citizen.

We combined this 5-item "naysaying" Low F scale with five items of the Positive F scale to form a new 10-item Balanced F scale. Thus, for this second population we had three different measures of authoritarianism: (*a*) the previously used (10-item) Positive F scale, (*b*) a new (10-item) Balanced F scale, and (*c*) a new (5-item) Low F scale. The correlational results indicated that we had been quite successful in constructing a "better" reversed F scale than has previously been reported. Our Naysaying Low F scale correlated highly negatively (−.70) with the Positive F scale. These "low" items also improved the Balanced F scale, as evident from its extremely high correlation (+.93) with the Positive F scale.

The correlations of these three F scales with the Agreement Response Scale are presented in Table 10.

None of the correlations of the F scales indicate any significant relationships to the ARS—further supporting our view that these dimensions are unrelated. We were puzzled that in this new sample the Positive F scale had no positive correlation to our new response tendency measure. Population differences may be involved, but other evidence, such as the slight positive relationships of "low" F scale items to the Agreement Response Scale, suggested that the new response scale was slightly "rotated" from the position expected from our original OAS criterion. But the ARS still seems quite adequate[3] as a "short" substitute measure; the general pattern of results reflects all the previous findings with the OAS measure.

Considering the correlations between response measures and authoritarian scales in our two samples, and also considering the orthogonality of the variables in the factor analysis of items, the best conclusion is that "pure" authoritarianism has little or no relationship to the agreeing response tendency.

TABLE 10

CORRELATION OF VARIOUS F SCALES WITH AGREEMENT RESPONSE SCALE

Scales	Correlation with ARS ($N = 54$)
Positive F scale	+.00
Balanced F scale (new)	−.04
Low F scale (new)	+.16

PRELIMINARY FORMULATION OF PERSONALITY DETERMINANTS FROM QUESTIONNAIRE RESULTS

From our investigation of the agreeing response tendency in the questionnaire realm, we have arrived at a fairly consistent picture of the variables that differentiate yeasayers from naysayers. Yeasayers seem to be "id-dominated" personalities, with little concern about or positive evaluation of an integrated control of their impulses. They say they express themselves freely and quickly. Their "psychological inertia" is very low, that is, very few secondary processes intervene as a screen between underlying wish and overt behavioral response. The yeasayers desire and actively search for emotional excitement in their environment. Novelty, movement, change, adventure—these provide the external stimuli for their emotionalism. They see the world as a stage where the main theme is an "acting out" of libidinal desires. In the same way, they seek and respond quickly to internal stimuli: their inner impulses are allowed ready expression. They have little tolerance for a delay of gratification: immediate reduction is sought for any tension that may have arisen. Thus, to both external and internal stimuli, the yeasayers' general attitude is one of *stimulus acceptance*, by which we mean a pervasive readiness to respond affirmatively or yield willingly to both inner and outer forces demanding expression.

The "disagreeing" naysayers have the opposite orientation. For them, impulses are seen as forces requiring control and perhaps, in some sense, as threats to general personal stability. The naysayer wants to maintain inner equilibrium; his secondary processes are extremely strong in this ego maintenance of the balance between emotionally impulsive and value-maintaining forces. We might describe this as a state of high psychological inertia—impulses undergo a series of delays, censorships, and transformations before they are permitted expression. Both internal and external stimuli that demand response are carefully scrutinized and evaluated: these forces appear as unwelcome intruders into a subjective world of "classical" balance. Thus, as opposed to the yeasayers, the naysayers' general attitude is one of *stimulus rejection*—a pervasive unwillingness to respond to impulsive or environmental forces.

---

[3] Additional validation of the Agreement Response Scale is provided by the correlation of +.61 with the sum of all 120 items in the second questionnaire battery.

This formulation of the dynamics of the response tendency was developed from an analysis of the empirical correlates of the Over-all Agreement Score. As such, the formulation suffers from all the limitations of psychological interpretation based on objective paper-and-pencil tests. Since the response tendency itself influences the objective test measurements by which we initially identified its nature, there were in this case additional reasons for further investigation. We felt that our study of the personality dynamics and determinants of agreeing response set would benefit from a clinical analysis of extreme yeasayers and naysayers. Therefore, a clinical study was designed to validate these theoretical formulations about the personality syndrome determining this general response set.

PART II: CLINICAL STUDY OF YEASAYERS AND NAYSAYERS

The research in Part I led us to the formulation that a personality syndrome may help to explain the psychological dynamics of agreeing response tendency. It seemed evident that for many persons the tendency to agree or disagree was actually a significant and functional part of their personality structure. The pattern of variables correlated with our two measures of the agreement tendency (OAS and ARS) suggested a set of hypotheses concerning the deeper dynamics and determinants of this pervasive response set. A clinical assessment of *S*s on the extreme ends of the dimension would allow us to test the main theoretical tenets of this "personality syndrome" view of the agreement response tendency.

Hypotheses about Personalty Differences

Specific predictions were made about differences between yeasayers and naysayers in seven major areas: (A) External vs. Internal Orientation; (B) Id-ego-superego Integration; (C) Impulse Control; (D) Hostility and Distrust; (E) Intellectual Orientation; (F) Secondary Process; and (G) Conflict Manifestations. These hypotheses were tested by a research design employing a sentence completion test and clinical interviews to be described below.

(A) *External vs. Internal Orientation.* The yeasayer is externally oriented with strong tendencies towards extraception and extraversion; the naysayer is internally oriented, intraceptive, and introverted.

A–1 Yeasayers have superficial, conventional identifications; naysayers' identifications are deeper, less conventional, and defined more by internal values.

A–2 Yeasayers are anti- or nonintraceptive; naysayers are intraceptive and psychologically minded.

A–3 Yeasayers are oriented outward, towards the group values; naysayers are oriented inward, towards their own values.

A–4 The self-concept of yeasayers is behaviorally defined and evaluated by conventional standards; that of the naysayers is subjectively defined and evaluated by personal standards.

A–5 Yeasayers have little concern with values and ideals, which if held at all are conventionally realistic and superficial; naysayers are continuously influenced by values which are felt as an integral part of themselves.

(B) *Id-ego-superego Integration.* The personality structure of yeasayers has a high degree of compartmentalization of id-ego-superego; naysayers have a more integrated personality with much stronger ego control.

B–1 The behavior of yeasayers is determined more by external factors in the immediate situation; naysayers are determined more by internal control and response to inner subjective reflections.

B–2 Yeasayers tend to externalize the source of their problems; naysayers tend to feel that they themselves are to blame for their problems.

B–3 Ego strength is low in yeasayers; high in naysayers.

B–4 Yeasayers tend to have weak, externalized superegos; naysayers are more guilt prone, having strong, internalized superegos.

B–5 Yeasayers show a low perseveration of mood and emotional cathexes; naysayers have a high perseveration of their feelings and attachments.

(C) *Impulse Control.* Yeasayers are able to express their emotions and impulses freely and openly without inhibition; naysayers tend to suppress, repress, and control their impulses by strong ego domination.

- C–1 Yeasayers express aggression and hostility immediately and openly; naysayers express hostility mostly in a covert way, while open aggression is extremely controlled.
- C–2 Yeasayers express affection, sex, tenderness more directly and easily than naysayers who are more constricted, inhibited, and repressed in the expression of libidinal impulses.
- C–3 Yeasayers have anal expulsive characteristics and respond to "anally colored" situations with typical expressive reactions; naysayers have anal retentive characteristics and show typical controlled reactions.
- C–4 Yeasayers were socialized by punitive, physical methods; naysayers were socialized by deprivation of love and rational persuasion.
- C–5 Yeasayers tend to have emotionally intense, "Dionysian" expressive behavior; naysayers stress an avoidance of strong feelings and favor a balanced, "Apollonian" form of expression.

(D) *Competition and Distrust.* The yeasayer tends to admit a distrust of others' altruism and to accept competition as a "realistic" part of life; naysayers have a "reactive" trust of others and try to minimize competition and conflict.

(E) *Intellectual Outlook.* The intellectual orientation of the yeasayers tends to be loose and emotional; naysayers are more consistently intellectual and have a high regard for rationality.

(F) *Secondary Process.* The yeasayer tends to have a minimum degree of secondary processes intervening between inner "stimulus" and manifest "response"; the naysayer has long secondary processes causing considerable delay in emotional responsiveness.

(G) *Conflict Manifestations.* Yeasayers show more evidence of conflict, ambivalence, and disparity between values and behavior than naysayers.

Method in Clinical Study

Subjects. From our original population of 61 *S*s, the 10 with the highest scores on the Over-all Agreement Score (yeasayers) and the 11 with the lowest scores (naysayers) were selected for the clinical study. The mean OAS score of the extreme yeasayers was 4.3 (range 4.2 to 4.5); the naysayers' mean was 3.6 (range 3.1 to 3.7).

Procedure. The yeasayers and naysayers were divided into two groups which were interviewed by one or another of the two *E*s. The OAS score of every *S* was carefully concealed from the *E*s. The interviews were oriented around a 55-item sentence completion test[4] which had been designed to test the major hypotheses outlined above. Several sentence "stems" in each hypothesis area were constructed to elicit responses that would differentiate yeasayers and naysayers along the lines of our theoretical formulations.

At the beginning of each individual interview, each *S* took the sentence completion test. The *E* then discussed the meaning of the *S*'s responses with him and tried to obtain information concerning the major areas of the authors' hypotheses. These interviews lasted from 2 to 4 hr. The *S*s were free to amplify, illustrate, or qualify their responses, and were frequently encouraged to digress when interesting themes appeared. The *E*s took extensive notes on the amplification of each response and also wrote short clinical evaluations of the *S*s in the different personality areas under study.

After all interviews were completed, both *E*s made independent ratings of sentence completion responses. Each *E* made 4 ratings of each of the 55 responses of the 21 *S*s on the basis of the following criteria:

1. Verbal content of response: Using *only* the written response, each item was rated on a 5-point scale (+2, +1, 0, −1, −2) which indicated the degree that the response corresponded to a "theoretical" yeasaying (+) or naysaying (−) orientation.
2. Degree of secondary process: On the basis of the written response alone, *E*s used the 5-point scale to rate the degree that each response was impulsive and immediate (+) showing a direct and uncensored reaction to the item; as contrasted with reflective responses (−) showing hesitation, qualification, rationalization, or other evidences of secondary processes.
3. Degree of conflict: Using both written responses and additional elaboration from the focused interview, each item was rated on a scale from 0 to +2 for the degree of conflict, ambivalence, and disparity between values and behavior.
4. Clinical assessment: Taking into account the elaboration and clinical material concerning each item, a clinical rating was made using the 5-point scale that assessed the degree of correspondence to a "typical" yeasaying or naysaying response.

A number of precautions were taken to reduce the possibility of interviewer bias in these ratings. After discussing the theoretical basis of the rating scheme, the *E*s did not discuss their ratings with each other. All identification was removed from the data. *E*s' ratings for each item were made separately; that is, all responses to an item were rated for all 21 *S*s before proceeding to another item. Thus, the *E*s did not know whom they were rating or what ratings the *S*'s other responses had been given.

As a first step in the analysis of the data, the Spearman-Brown reliability of the ratings was studied. Table 11 presents the reliability coefficients for the sum of *E*s' clinical ratings for the variables in each of the major hypothesis areas.

For the sums of subareas and for individual items, reliability coefficients were somewhat lower; the average for subareas was +.69, for individual items +.61. On the whole interrater reliability seemed very satisfactory; therefore, ratings of the two *E*s were pooled to provide the basic data for the authors' subsequent analysis.

---

[4] The Sentence Completion Test used in this study has been deposited with the American Documentation Institute. Order Document No. 6134 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 11

RELIABILITY COEFFICIENTS FOR INTERVIEW RATINGS

Major Hypothesis Area	Verbal Content Ratings ($N = 21$)	Clinical Assessment Ratings ($N = 21$)
A—External-Internal Orientation	.57	.68
B—Id-ego-superego Integration	.80	.86
C—Impulse Control	.95	.89
D—Competition and Distrust	.47	.60
E—Intellectual Outlook	.56	.83
F—Secondary Process (sum for all items)	.76	—
G—Conflict Manifestations (sum for all items)	—	.65

Results of the Clinical Analysis

The main analysis of the data was a series of t tests on the difference of means between the groups of extreme yeasayers and naysayers. The tests were performed for each of the major hypothesis areas, and for the subareas that were stated in the previous section. In Table 12, the results of the t tests are given. Each hypothesis area is stated in a yeasaying form, so that a positive t value indicates that extreme yeasayers have a higher mean than extreme naysayers. After each hypothesis, we have indicated in parentheses the number of sentence completion items included by design in testing that particular hypothesis. The results for ratings using only the verbal content of the sentence completions, and for ratings that include clinical assessment are tabulated separately.

In the five major areas of personality—(A) External-Internal Orientation, (B) Id-ego-superego Integration, (C) Impulse Control, (D) Competition and Distrust, and (E) Intellectual Outlook—for which we had formulated systematic hypotheses, all results for the clinical assessment ratings were significant at the .01 level or greater. The analysis of the verbal content ratings shows the same pattern at somewhat lower levels of significance. For the subarea hypotheses, 9 out of 15 tests on the clinical ratings were significant at the .05 level or greater, with particularly high t values for Anal Expulsive Expression. These results indicate that ratings that take into account the additional material from the interview (Clinical ratings) differentiate the extreme groups better than the ratings based only on the written answers to the sentence completion items (Verbal Content ratings).

The hypothesis that yeasayers would show less evidence of Secondary Processes than naysayers was supported at only the .10 level of significance ($t = +1.62$). However, additional evidence from the data supports the hypothesis: of the 55 items involved in these ratings, 45 are in the predicted direction, giving a highly significant chi square of 21.9.

Despite the positive nature of most results, some hypotheses were not confirmed. The supposition that Conflict Manifestations would be greater for yeasayers was not at all supported ($t = -0.65$). The expected Conventional Identification of yeasayers was not indicated by either the Verbal Content ratings ($t = +0.40$) nor by the Clinical ratings ($t = +0.94$). In the same way, the hypothesized Externalization of Problems, Weak Superego, and Punitive Physical Socialization, as characteristics of yeasayers, were not significantly supported by the data.

Discussion of Clinical Findings

The clinical study clearly supports the main lines of the hypothesized personality syndrome behind the agreeing response tendency. The analysis of interview ratings showed that yeasayers and naysayers were significantly differentiated in the areas of External-Internal Orientation, Id-ego-superego Integration, Impulse Control, Competition and Distrust, Intellectual Outlook, and Secondary Processes. These personality differences appear in many diverse areas of functioning. The generality of the manifestations supports our contention that response set is based on an underlying dynamic syndrome. The nature of the syndrome makes it possible to explain a marked response set to the "stimuli" of questionnaire items as a functional consequence of inner forces in the responding individual. Questionnaire items are another kind of stimuli demanding response. As such they bring forth typical reactions of immediate acceptance from yeasayers and of inhibited resistance from naysayers.

External Orientation. Many facets of the difference in External Orientation between yeasayers and naysayers were made evident

TABLE 12

THE *t* TESTS OF DIFFERENCES BETWEEN EXTREME YEASAYERS AND NAYSAYERS

Hypothesis Area	Verbal Content Ratings	Clinical Ratings
A—External Orientation (10)	+1.89*	+3.89**
A-1 Conventional Identification (2)	+0.40	+0.94
A-2 Anti-introception (2)	+1.86*	+2.33*
A-3 Outward Group Orientation (2)	+1.67	+4.21**
A-4 Conventional Self-concept (2)	−0.10	+1.96*
A-5 Superficial Values and Ideals (2)	+3.19**	+2.94**
B—Lack of Id-ego-superego Integration (10)	+1.48	+2.52**
B-1 Situational Determinism of Behavior (2)	+1.10	+2.27*
B-2 Externalization of Problems (2)	−0.14	+1.28
B-3 Low Ego Strength (2)	+1.77*	+1.55
B-4 Weak Superego (2)	−0.19	+0.61
B-5 Rapid Changes in Emotions (2)	+1.31	+1.70
C—Impulse Expression (25)	+2.35*	+3.03**
C-1 Overt Expression of Hostility (5)	+2.36*	+3.27**
C-2 Overt Expression of Affection (5)	+1.74*	+2.27*
C-3 Anal Expulsive Expression (5)	+2.69**	+3.28**
C-4 Punitive Physical Socialization (5)	+0.57	+1.09
C-5 Intense "Dionysian" Expressiveness (5)	+1.02	+2.47**
D—Acceptance of Competition and Distrust (5)	+3.01**	+3.30**
E—Lack of Intellectual Outlook (5)	+1.74*	+3.22**
F—Low Secondary Processes (55) (verbal content alone)	+1.62	—
G—Conflict Manifestations (55) (clinical rating alone)	—	−0.65

* Pooled *t* tests were used for testing the significance of the differences in means between ratings for (10) yeasayers and (11) naysayers. For 19 *df* the .05 level of significance for a one-tailed test is 1.73.

** The .01 level of significance is 2.43.

by observation of their general manner and behavior during the interview situation. It was apparent from our clinical reports that yeasayers were much more extraverted in their social orientations. They were usually more voluble, spoke more easily about themselves (though often with little depth), took suggestions readily, reacted in more obvious and overt ways to their moods, appeared less complex psychologically, tended to wander from the subject at hand, and made themselves thoroughly at ease with the interviewer. Naysayers, on the other hand, were characteristically more introverted. They were often shy and always reserved about themselves. Throughout the interview, they required more questioning and prodding; they felt easily criticized, weighed and considered each answer, and tended to express themselves in indirect and complicated ways—perhaps a consequence of the fact that for many of them introspection and self-analysis seemed virtually a full time occupation. By their introverted behavior and general lack of social spontaneity, the naysayers tended to create a task oriented atmosphere and usually initiated and maintained an intellectual relationship with the interviewers.

Impulse Control. Our hypotheses concerning Impulse Control, which had been fully supported by the empirical analysis of Clinical ratings, were verified more indirectly by our observations during the interview. In discussing impulsivity and the expression of feeling, yeasayers and naysayers were especially different. Yeasayers talked with enthusiasm about their impulsivity and their love of emotionally exciting and novel situations. Sexuality and aggressive feelings were discussed quite openly, often with exhibitionistic exaggeration. Yeasayers sometimes expressed the fear that they were excessively uncontrolled and attributed their interpersonal difficulties to impulsiveness, hastiness, or unthinking rashness. Naysayers, on the other hand, were often defensive and embarrassed about emotional impulsivity and their manner of expressing inner feelings. They were frequently worried about their shyness and lack of social poise. They seemed concerned about their excessive emotional constraint and lack of spontaneity. Yet beneath this surface of emotional constriction, there appeared to be a depth of strong feelings which could only come out in close personal relationships.

Intellectual Orientation. A striking difference in the Intellectual Orientation of naysayers and yeasayers was apparent throughout the interview. Naysayers had a pervasively rational and intellectual approach to whatever topic was discussed, while yeasayers were typically less conceptual, viewing the world in terms of emotional reactions and concrete instances. In exploring this area, we asked the subjects about their attitude towards questionnaires. Yeasayers and naysayers described how they approached the task of answering Likert scale items in consistently different ways. The yeasayer responded to the "surface" of the items' meaning, did not reflect on their possible full implications, and thus gave impulsive, immediate answers. This pattern is consistent with the yeasayer's preference for blunt, straightforward, and extreme opinions, which reflect his emotional feelings at the moment. In quite an opposite manner, the naysayer analyzes and dissects each item, considering it from several points of view and attempting to be "logical" in his responses to the highly "interpreted" statements of the questionnaires. As was noted in our discussion of item tone, there seems to be an affinity of the yeasayer and naysayer for items phrased in his characteristic style of thought. Yeasayers are themselves enthusiastic, colloquial, and given to extreme opinions. Naysayers are guarded, qualified, and somewhat negative in their intellectual orientation. When the phraseology of a questionnaire item approaches one of these patterns, it presents a differential attraction to these two types of individuals. These differences in verbal style and intellectual approach contribute to the opposing patterns of the extraverted, fun loving, emotional yeasayer and the introverted, serious-minded, and rational naysayer.

Personality Integration and Conflict. In the area of Id-ego-superego Integration our hypotheses were supported on an over-all basis, but some of the subhypotheses were below normal significance levels, in particular, Externalization of Problems and Weak Superego. A similar error had been made in our hypothesis about personality conflict. The ratings of Conflict Manifestations had failed to show the evidence of ambivalence, denial, and dissociation that we had expected to differentiate yeasayers and naysayers. In attempting to explain these negative findings, we noted that all of these hypotheses had been derived from the questionnaire results that portrayed yeasayers as anxious, psychasthenic, schizoid, unstable, and generally maladjusted. Our clinical study indicates that the implications of these scale correlations must be re-evaluated. From the interview, it was apparent that yeasayers were more *able and willing to admit* their "unacceptable" impulsive feelings; whereas naysayers were unusually susceptible to denial mechanisms and a general inhibition or repression of libidinal and aggressive desires. In discussing competitive situations and aggression, the naysayers were often unaware of hostilities which seemed obvious to the interviewers. Their "reactive" trust had, at times, the characteristics of repressed hostility and resentment. On the other hand, the yeasayers were in much closer contact with their unacceptable impulses and had fewer inhibitions about revealing them than the naysayers.

It is thus probable that the yeasayers' higher scores on "maladjustment" scales, which had suggested our hypothesis of dissociation and conflict, are chiefly attributable to their greater frankness and willingness to admit impulsive feelings and ambivalences. Clinical results also suggest that the naysayers may be presenting a "mature" picture colored by strong denial mechanisms. Two of our naysaying *S*s later in the year developed rather severe neurotic disturbances requiring therapeutic treatment; the questionnaire results gave woefully inadequate indications of these trends. We would now reformulate our hypotheses about these aspects of dissociative tendencies to predict the type, but not the degree of conflict or maladjustment. The agreeing and disagreeing dimension seems indicative only of the central area where psychogenic deviation may occur: yeasayers' problems center around inadequate controls, while naysayers' are focused on excessive inhibition. To the extent that yeasayers lack personality integration, they deviate in the direction of "id domination" associated with an externalized superego and ineffectual controls. On the other hand, the naysayer who becomes maladjusted should tend towards "superego domination," frequently maintained by emotional dissociation and repressive controls.

Conventionality. In the area of superego strength and conventional values, another set of hypotheses was not fully supported. Besides the Weak Superego hypothesis already mentioned, two others, Conventional Identification and Punitive Physical Socialization, indicated an error in prediction, based in part on the premise of dissociation discussed above, but also on the assumption of greater "conventionality" for the yeasayer. This latter assumption seems to be in error. Contrary to prediction, many naysayers had conventional religious ideologies that were strongly internalized; naysayers were also more conventional in many areas of social behavior and attitude than the more "liberated" and exhibitionistic yeasayers. Interview material also indicated that the type of identification with the socializer was more relevant to the response set dimension than the particular method of socialization. The parents remain external figures to the yeasayer; whereas the naysayer identifies more completely with them. Yeasayers seemed freer to express their impulses with little concern for interpersonal consequences. Yet these same yeasayers had "strong" but rigid superegos which had little influence on their everyday behavior, except in the sense that this "control" had to be opposed and subverted. They experienced guilt as something external to their central selves. Naysayers, on the other hand, had superegos firmly integrated into their ego ideals. In our original formulations, we had erred in stressing superego strength, instead of the degree of value internalization. It is the degree of internalization and not the degree of conventionality that determines the differentiating influence of values on the behavior of yeasayers and naysayers.

Awareness of Response Set. Toward the end of each interview, the topic of response set was brought up explicitly. The *S*s were reminded of the questionnaires they had taken earlier in the year, and were asked whether they had, in general, agreed or disagreed more with the items. More than three-fourths of the *S*s characterized their response set accurately; one was inaccurate and the rest had no recollection. Most *S*s reported some awareness of response bias as they filled out questionnaires. Several commented that they had consciously compensated for their tendency to mark predominantly in one direction or the other, but as is evident from the basis of their selection as extreme yeasayers and naysayers, in all cases the original response set had prevailed.

GENERAL INTERPRETATION AND CONCLUSION

Statistical Implications of Response Set

Our research originated with a realization that agreeing response set was of far-reaching importance in the realm of objective questionnaire tests. The usual assumption of psychological test construction is that agreement or disagreement with a scale of items measures primarily the variable for which the test was designed. But many of the statistical techniques used in scale development result in a contamination of the final scale with an unwanted component of response set. No matter how a scale is developed, it is virtually impossible to eliminate the effect of agreement bias in an *S*'s answer pattern. As a consequence questionnaire items seem almost "possessed" by a tendency to correlate positively with one another regardless of content. The *S* is, in part, responding to the "meaning" of the item, but also, in part, to the answer scale associated with the item. Although this answer scale is a necessary part of the measurement technique,[5] it also evokes the agreement response tendency that subsequently becomes inextricably interwoven into the item score. It is quite possible that many factor analyses of item domains have revealed a first centroid factor which might best be explained as a manifestation of response set.

---

[5] Some measurement techniques such as Thurstone's Paired-Comparison Method, Guttman Scale checking, forced choice alternative responses may avoid some of the intrinsic effects of the agreement tendency. But even here the response bias may appear in subtle ways, such as a tendency to check every item or the most extreme item.

This same phenomenon is evident with multi-item Positive scales which tend to have suspiciously high positive intercorrelations with one another. The diagnosis is again "response set," operating all the more powerfully when several items are summed in the same scoring direction to measure an intended variable. These scale scores also measure in varying and unknown degrees a component of the agreement response tendency. It is this "common factor" of response bias which often causes the "operational" measurements of variables to correlate positively, regardless of the degree of psychological "opposition" of the measured concepts. The sum of any group of Likert scale items scored positively will always be a somewhat valid, though perhaps unintended, scale for response set. These Positive scales will show the effects of this additional component, partly by being prone to positive interrelationships with other Positive scales, and partly by reflecting in their pattern of highest correlations some of the psychological aspects of the agreement personality dimension. Since all intercorrelations with scales of this type are pushed in a positive direction, correlational results must be interpreted with this response variable in mind. It also follows that significant negative correlations will appear only when there are extremely strong factors operating against the positive influence of response set. An indication of this bias was apparent in the questionnaire battery where nearly 90% of all items correlated positively with our measure of agreement tendency.

Balanced scales. In our attempt to study the response tendency in the questionnaire realm, we remained acutely aware of these difficulties. In trying to identify the properties of the response variable, we had occasion to make use of "balanced" scales, containing an equal number of items scored positively and negatively. This type of scale is a useful method of cancelling out at its source the contaminating influence of agreement response set on sum scores, and thus provides at least a statistical solution to some of the difficulties raised by this biasing tendency.

Response Set as a Personality Variable

Our primary interest in this study has been in the fundamental psychological determinants of agreeing response set within the responders themselves. Basing our approach on the conviction that response set is a reflection of personality dynamics, we set out to explore its correlates in the realm of objective personality tests. An analysis of these relationships suggested that the best single characterization of the traits associated with agreeing response set is *Stimulus Acceptance* vs. *Stimulus Rejection*. The yeasayer accepts stimuli both by admitting them to consciousness without censorship, alteration, or assimilation and by agreeing with, acting out, and otherwise yielding to the pressures of stimuli exerted on him. The naysayer tends to reject these same stimuli, which may be either internal (impulses, needs, whims, sudden moods) or external (suggestions, provocations, coercions, excitement, novelty). The most obvious manifestation of the general acceptance of stimuli takes the form of "agreeing" with, marking Yes, or checking True to a wide variety of otherwise unrelated items and statements.

Our interview study brought to light many facets of inner dynamics which permit us to go beyond a simple trait explanation of the agreeing and disagreeing response tendencies. In terms of personality structure, the most salient difference between yeasayers and naysayers is in their mode of *ego functioning*. Our results suggest that yeasayers have relatively "passive" (releasing) egos, as contrasted with the more "active" (controlling) egos of the naysayers. This difference appeared most clearly in the contrasting strength of secondary processes in their personality structure. For naysayers, the secondary processes of delaying, censoring, rationalizing, sublimating, and integrating were far more prominent than in the less active egos of yeasayers who portrayed themselves as unreflective, quick to act, easily influenced, and unable to tolerate delays in gratification.

These differences in ego functioning might best be characterized in terms of the *psychological inertia* of secondary processes. High inertia characterizes the active egos of naysayers who have a high resistance to movement and change, a slowness in their responses, a strong tendency toward equilibrium in their inner world, and an ego domination that frequently appears in excessive control and inhibition. On the other hand, yeasayers are low in psychological inertia, reacting quickly and overtly to emotional pressures without the "dampening" effect of inhibiting and delaying mechanisms.

Theory of Developmental Dynamics of Response Set

To achieve a deeper understanding of this pattern of personality factors, it seems necessary to formulate the developmental dynamics of agreeing response set. The data in both parts of the study suggest that the contrasting personality patterns of yeasayers and naysayers have their developmental origins in the differential reactions to what are often called anal socialization problems. Such an interpretation, while not directly tested by this analysis, nonetheless helps to integrate the results of our study.

Psychoanalytic theory (Freud, 1908, 1916) emphasizes that the resolution of early problems centering around the control of elimination underlies later patterns of impulse control. The child's parents exert pressure on him to control his own evacuation (and other antisocial, value-violating impulses) with the hope that he will internalize their commands and enable them to relax their vigilance. One type of reaction to such parental pressure is *not to internalize* the parental edicts, but to retain the need for external control and guidance, especially in the area of impulse control. This "solution" involves the child in a continual conflict between his desires and those of the restraining authorities, but at the same time allows him (both as child and adult) the pleasure of expressing impulses which are supposed to be suppressed. The pleasure is often a double one; partly a narcissistic gratification and partly the attention-getting satisfaction of annoying external authorities. This reaction could be called the *anal expressive* or expulsive resolution of early socialization pressures, and is associated with ineffectual control of impulses. The second type of resolution involves *a complete internalization* of parental control functions, perhaps related to an unusually strong identification with the parents. In addition to the superego development stemming from this internalization, the ego of the child gains autonomy and power. In a sense, he becomes his own parent by transferring to his own personality the previously external power of the parents to control and reward, thereby gaining self-control and active mastery of his instinctual demands (Erikson, 1950, pp. 76–81). He thus gains his freedom from external restraint, but at the price of severe internal control. When acting like the controlling parent, the child is rewarded from within for suppressing his own "childish" impulses. With these developments comes the desire and ability to delay gratification, the main component of active ego functioning. Since the parent-like ego of the child now has the ability to "retain" or suppress impulse expression, this resolution of control problems may be termed anal retentive or *anal suppressive.*

Our results suggest that yeasaying and naysaying are closely related to these two opposite reactions to parental pressure for control of antisocial behavior. Yeasayers show many manifestations of the anal "expressive" type of resolution that involves the lack of internalization of impulse control. Their original uncontrolled expressiveness is only slightly transmuted into a conscious preoccupation with "anally tinged" gratifications and expressive exhibitionism. Not having strong inner controls, they express their impulsive desires with little or no guilt. Evidences of these anal concerns among yeasayers are especially prominent in our questionnaire studies. The Anal Preoccupation factorial scale, which correlated +.75 with the OAS, contains items expressed in implicitly anal metaphors, such as: "really splurging on something," "sitting surrounded by your possessions," "sitting in the seats of the mighty," "smelling offensive odors," "saying dirty things," "spending money freely," and "exploring dirty slums and alleys." The Anal Submission factorial scale, which correlated +.67 with the OAS, contains a set of items expressing a general need for external control: "The college student needs advice and guidance from someone wiser than himself," "I often wish that someone could make my decisions for me," etc. In these items, we see the anal pattern in all of its manifestations; on the one hand, the craving for excitement and the continual search for and indulgence in the pleasures of impulse expression that remains unrestrained by the ego, and on the other hand, the need for external control and guidance, closely coupled with resentment of this same control.

The naysayers' retentive or suppressive resolution of anal problems shows many manifestations of their internalization of parental demands for control and of their subsequent aversion for "anally expressive" impulses. On questionnaire items of the "anal" variety discussed above, the naysayer consistently disagrees. He is emphatically *not* interested in "splurging," "messing around," or "blowing up"—he has learned his early lessons well, and these impulses have ceased consciously to trouble him. His identification (sometimes overidentification) with the antianal concerns of his parents is also shown by his rigidly maintained "suppressive" control of emotional responsiveness.

The focused clinical study considerably expanded this formulation of the role of anality in the personality dynamics of naysayers. Their identification with the socializing pressures of their parents appears not only in their dislike of anal concerns but also in their avoidance of "childish" expressions of impulses. Having taken upon themselves the full burden of what was once external control, they no longer feel any need for external advice, guidance, or counsel. The necessity of friction and aggressiveness in competitive situations is strongly denied, and replaced consciously by reactive trust and tolerance for others. In interviews, naysayers are extremely reluctant to admit to any hostility toward their close friends, and overt aggressiveness is avoided at all costs. They avoid all other situations that provoke emotional expressiveness or an abandonment of ego controls. Much energy is expended in the effort to maintain a calm equilibrium. Lacking the youthful exuberance or carefree spontaneity of the yeasayer, the naysayer stresses sublimated satisfactions, stability, and self-control, damping down potentially disruptive emotions and avoiding situations which might disturb his balance. With little visible affect and great surface calm, he turns his attention toward introspection and his complex inner life. But, in extreme form, this overcontrolled personality pattern suggests that something may be amiss beneath the defiantly calm surface. Avenues toward libidinal gratification are blocked and the transmutation of aggression into reactive trust and tolerance may, in some cases, cause extensive parataxis and constriction of interpersonal relationships.

In conclusion, we have tried to integrate the findings of our study of response set by presenting a theory of the developmental dynamics of "yeasaying" and "naysaying." This theory is speculative, but the consistent evidence from both the questionnaire data and clinical results supports our contention that agreement response tendency is ultimately related to a meaningful personality syndrome which is consistent with psychoanalytic accounts of the resolution of problems of impulse control.

SUMMARY

This study began with a realization of the importance of the agreeing response set in the field of personality measurements. It was evident that an agreeing-disagreeing tendency was permeating the answer patterns of our subjects to a series of questionnaire items. Our first step was to develop an appropriate measure of this response set which would be essentially independent of any specific content. By increasing the number and heterogeneity of items, and by "cancelling out" positive and negative statements of each variable, we obtained a 360-item Over-all Agreement Score (OAS) that met this criterion.

This measure of response bias showed considerable internal reliability, consistency over time, and generality over tests. The agreeing tendency was highly related to "true-saying" and "yes-saying" on other standard tests. We then investigated the correlates of the OAS to the variables of several objective test domains. A cluster of scales characterized the positive end of the agreeing tendency:

Impulsivity, Dependency, Anxiety, Mania, Anal Preoccupation, and Anal Resentment. The opposite end, the "disagreeing tendency," was defined by such scales as: Ego strength, Stability, Responsibility, Tolerance, and Impulse Control. When a factor analysis of individual items was performed, a major dimension of Stimulus Acceptance vs. Stimulus Rejection emerged as central to the agreeing-disagreeing response set. Personality test results also indicated pervasive differences in ego functioning, particularly as regards the high vs. low psychological inertia of secondary processes.

In the course of our study with objective questionnaires, we explored the technique of constructing Balanced scales and noted their effectiveness in eliminating the influence of the agreement response set on scale scores. By using an equal number of positively and negatively scored items, the agreeing response set can be balanced out at its source. In this manner, we could demonstrate the lack of any strong relationship of authoritarianism to agreement response set. The independence of response set and authoritarianism was supported by a factor analytic study where these variables emerged as orthogonal dimensions. In our further research with questionnaire items, we noted that "item tone" was a significant influence in attracting or repelling agreement. By using this knowledge, we constructed a "Naysaying" Low F scale that showed the expected characteristics of a nonauthoritarian measure.

Having discovered a fairly definite psychological pattern of objective scales that were empirically related to the Over-all Agreement Score, we proceeded to a clinical analysis of extreme *S*s at opposite ends of this agreeing response tendency. A series of hypotheses were formulated about the personality differences between "yeasayers" and "naysayers." Using a clinical interview structured around a sentence completion test, we confirmed the main aspects of our personality formulation of the agreement response tendency. Yeasayers were shown to be individuals with weak ego controls, who accept impulses without reservation, and who "agree" and easily respond to stimuli exerted on them. The naysayer inhibits and suppresses his impulses, in many ways rejecting all emotional stimuli impinging on him.

The clinical study suggested a possible developmental explanation of the deeper determinants of the yeasayer's and naysayer's personality structure. During their early socialization period, they tend towards opposite resolutions of the "anal" problems around control of impulses. The passive and receptive egos of yeasayers seem to be an outgrowth of their lack of internalization of parental control. While maintaining the need for external control, yeasayers can freely indulge in impulse gratification. On the other extreme, the active, guiding egos of the naysayers apparently result from an internalization of the controlling functions of the parental figures they had strongly identified with during childhood. The egos of naysayers take over the controlling functions of the parents, and the suppression of impulses is subsequently self-maintained and self-rewarded.

In conclusion, we feel that this integrated study, though limited by the size and nature of our population, has demonstrated both the far-reaching importance of response set in the area of psychological tests and the major proposition that the agreeing response tendency is based on a central personality syndrome.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BARNES, E. H. Response bias in the MMPI. *J. consult. Psychol.*, 1956, **20**, 371–374.

BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

BERDIE, R. F. Likes, dislikes, and vocational interests. *J. appl. Psychol.*, 1943, **27**, 180–189.

CATTELL, R. B., SAUNDERS, D. R., & STICE, G. E. *The 16 P. F. questionnaire.* (rev. ed.) Champaign, Ill.: Inst. Pers. and Ability Testing, 1957.

CHRISTIE, R., HAVEL, J., & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

COUCH, A. S., & BALES, R. F. *Factor analysis of value profile.* 1955. (Mimeo.)

CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, **6**, 616–623.

CRONBACH, L. J. *Essentials of psychological testing.* New York: Harper, 1949.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & T. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958.

DAVIDS, A. Alienation, social apperception, and ego structure. *J. consult. Psychol.*, 1955, **19**, 21–27.

ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.

FREUD, S. Character and anal eroticism (1908). In *Collected papers*. Vol. II. London: Hogarth Press, 1946.

FREUD, S. On the transformation of instincts with special reference to anal eroticism (1916). *Collected papers*. Vol. II. London: Hogarth Press, 1946.

FRICKE, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.

HOLLAND, J. L. A personality inventory employing occupational titles. *J. appl. Psychol.*, 1958, **42**, 336–342.

JACKSON, D. N., & Messick, S. J. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.

KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, **23**, 226–236.

LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215–221.

LORGE, I. Gen-like: Halo or reality. *Psychol. Bull.*, 1937, **34**, 545–546. (Abstract)

ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.*, 1956, **70**, (18, Whole No. 425).

ROULON, P. J. *A statistical study of the strong vocational interest blank*. Unpublished doctoral dissertation, Univer. of Minnesota, 1931.

THOMSON, G. H. *The factorial analysis of human ability*. London: Univer. London Press, 1939.

THURSTONE, L. L. *Thurstone temperamental schedule*. Chicago: Science Research Associates, 1949.

(Received February 5, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

ACHIEVEMENT AND NURTURANCE IN TRIADS VARYING IN POWER DISTRIBUTION[1]

MARILYN V. CHANEY AND W. EDGAR VINACKE

University of Hawaii

THE following study is part of a program of research concerned with intragroup relations (coalition-formation or non-formation) that develop in a competitive situation where winning depends upon establishing an appropriate bargain with another member of the group. The point of departure was provided by a series of hypotheses suggested by Caplow (1956), who based his analysis of power situations upon work by Georg Simmel (Wolff, 1950; Mills; 1954, 1954) and others. Simmel held that any group of three tends to become a coalition of two against one, but that these alliances may shift from pair to pair, depending upon many factors, including the immediate issue and the personality characteristics of members.

Caplow proposed that coalition formation should be related to the internal power differentiation of the triad. A simplified statement of his propositions for six power patterns is as follows:

Type I (1-1-1). When all are equal, all possible pairs are equally likely to form a coalition.

Type II (3-2-2). When one member is stronger than the other two (who are equal), but in combination exceed the first, the two weaker members will usually ally against the stronger.

Type III (1-2-2). When one is weaker than the other two (who are equal), but the weaker in combination with either of the others exceeds the third, this pairing will occur more often than that of the two stronger members.

Type IV (3-1-1). When one is stronger than the other two combined (who are equal) no coalition will ordinarily occur.

Type V (4-3-2). When all are different in strength, but any pair exceeds the third person, then the two weaker will tend to ally against the strongest.

Type VI (4-2-1). When all are different in strength, but one is stronger than the other two combined, no coalition will very often occur.

Vinacke and Arkoff (1957) tested Caplow's hypotheses in a simple game situation (to be here described). Their results sustained all the hypotheses. The general strategy displayed was interpreted as depending upon the *perception of initial strength*, in contrast to an alternative strategy of *rational analysis of outcome*, derived from the theory of games (see Vinacke & Arkoff, 1957). The former strategy has characterized both male and female, as well as mixed-sex, triads, playing under several incentive conditions.

The question now arises concerning the effect of personality variables upon play when power pattern varies. A number of studies, employing three-person groups, are suggestive in this regard. Mills (1954) found that differences in social status have little effect on coalition formation, but that need for self-enhancement is more important in high- than in low-status members. Chapman and Campbell (1957), varying the authoritarianism of members, found no significant superiority for any combination. DeCharms (1957), in contrast, showed that threat-oriented affiliation results in high productivity in competitive tasks, but low productivity in cooperative tasks. In a study by Hoffman, Festinger, and Lawrence (1954), the importance of the task and the degree of comparability (perception of others as peers or non-peers) both proved to affect strategy in a game situation. Thus, weaker members more often allied against a person with a large initial advantage when the task was of little importance and the strong member was thought to be superior to them.

In order to manipulate motivational conditions, we selected two of Murray's "manifest" needs (1938) most obviously likely to influence play in a competitive game. "Achievement" has been described as desire to overcome obstacles, exercise power, and to strive to do something difficult as well and as quickly as possible. It is said to be characterized in part by enjoyment of competition and determination to win. "Nurturance," in contrast, is defined as need to nourish, aid, or protect the helpless, to express sympathy, and to be associated with traits of generosity, sympathy, and tolerance towards others.

---

[1] This article is based on the senior author's masters' thesis deposited in the University of Hawaii Gregg M. Sinclair Library. The analysis hereinafter presented, however, differs from the original, which used games as units rather than pairs and individuals.

The problem of this experiment is to compare the behavior of subjects (*S*s) high in achievement with *S*s high in nurturance in a competitive three-person game, with the third member serving as a control. It was hypothesized that the former would display a pattern of behavior reflecting desire and effort to win, the latter a more accepting and less competitive style of play.

METHOD

Subjects

Twenty male triads were recruited from the introductory psychology course. All students were administered the Edwards Personal Preference Schedule (Edwards, 1954), a test that purports to measure motivational tendencies in accordance with Murray's formulations. Among the 15 variables assessed by the test are achievement and nurturance (the correlation between them, as reported by Edwards, is $r = -.30$). *S*s high in achievement were selected from those falling at or above the 74th percentile on that scale and at or below the 24th percentile on nurturance. For nurturance, *S*s were chosen from those who fell at or above the 70th percentile on that scale and at or below the 24th percentile on achievement. Control *S*s were taken from those between the 50th and 58th percentiles on achievement and the 37th to 63rd percentiles on nurturance. Each triad was composed, in this fashion, of an *S* high in achievement, an *S* high in nurturance, and a control. The characteristics of the triads may be summarized, as follows:

	Achievement Score		Nurturance Score	
	M	*SD*	*M*	*SD*
Ach *S*s	20.75	1.99	9.25	2.57
Con *S*s	15.25	1.52	13.25	1.65
Nur *S*s	9.70	2.36	21.15	2.64

Game Situation

A simple game was devised by pasting numbers consecutively on the outside spaces of an ordinary parchisi board. To the colored counteres (red, green, blue, and yellow) were affixed on one side numbers corresponding to the requisite weights that were to determine the strength of the players in the game. The power types used in this experiment were the Types I, II, V, and VI, defined above, and referred to below as Type I (1-1-1), Type II (3-2-2), Type III (4-3-2), and Type IV (4-2-1). Each triad played through 12 games in three series of each type, with position of the type in each series randomized. The three required counters were presented to *S*s, with weights concealed, in a hopper, from which they drew in turn (order rotated among *S*s). This technique did not guarantee an equal occurrence of each weight for each *S*, but did enhance interest in the game, a condition deemed essential. An arbitrary "prize" of 100 points could be won, all or in part, in every game.

Procedure

After the triad was assembled (without mention of the motive test previously administered in the lecture class), *E* explained that they would play a game and presented each member with a copy of the following instructions:

> This game is a contest among three players. Before the game begins, each player draws a marker out of the hopper. This marker indicates the strength that the player is to have during the game. When the referee throws the die, each player receives one space for each spot shown, and the player moves forward the number of spaces to which his strength entitles him. For example, if there is a "2" on the marker and the die shows "3," the player will move forward six spaces.
>
> The game is won by the first player to encircle the board. This player is awarded 100 points except that in the case of ties, the winning players will divide the prize equally among themselves.
>
> At any time during the game, any player, in return for a promise of a specified portion of the prize, may form an alliance with any other player. In this case, the allies immediately proceed to the position equal to their combined acquired spaces, and in the future throws they move together according to the total of their combined strengths.
>
> Any player may concede defeat when his position seems hopeless.

After the instructions had been clarified, as necessary, the first draw was conducted. Thereafter players were left to their own devices, except to assure that they played according to the rules. *E* remained an impartial referee throughout the play. He kept a record of weights drawn, progress of play, alliances offered and formed, and other relevant information.

All triads soon discovered that actual play is not necessary once coalitions are agreed upon, or, in Type IV, once the draw is completed.

RESULTS

The data concerning coalitions are presented in Table 1. Two points of explanation

TABLE 1

COALITIONS AMONG THREE MOTIVE-CONDITIONS IN TERMS OF OPPORTUNITY
(N = 20 Triads)

% Opportunity	Type I			Type II (3-2-2)						Type III (4-3-2)								
	(1-1-1)			Strong (32)			Weak (22)			Strong (43)			Strong (42)			Weak (32)		
	AC f	AN* f	CN f	AC f	AN f	CN f	AC f	AN f	CN f									
51 or more	4	0	9	0	1	0	6	9	14	1	2	2	1	1	1	8	5	9
33 to 50	8	10	7	4	4	2	0	3	2	2	1	2	1	4	1	2	2	1
1 to 32	5	8	2	1	1	0	0	0	0	1	1	1	1	0	1	0	1	0
0	3	2	2	15	13	14	2	2	2	10	8	11	8	12	11	2	4	3
	20	20	20	20	19	16	8	14	18	14	12	16	11	17	14	12	12	13
M%	36	26	50	9	14	5	75	70	78	16	22	19	15	16	11	68	48	67
(Triple Coalitions)	(28)			(5)						(6)								

Note—Explanations of entries: A, C, N refer to achievement, control, and nurturance Ss; each frequency refers to a pair (or S) in a single triad; the dashed line shows the division employed in computing chi square. Triple alliances are prorated among pairs.

* AN vs. CN: $\chi^2 = 3.96$, $n = 1$, $P < 0.5$. All other testable comparisons not significant.

TABLE 2

PARTICIPATION IN COALITIONS BY THREE MOTIVATION CONDITIONS IN RELATION TO OPPORTUNITY
(N = 20 Each)

% Opportunity	Type I			Type II (3 -2-2)						Type III (4-3-2)								
	(1-1-1)			Strong (as 3)			Weak (as 2)			Strong (as 4)			Middle (as 3)			Weak (as 2)		
	A* f	C f	N f	A f	C f	N f	A f	C f	N f									
51 or more	4	15	10	2	2	2	11	13	17	3	4	2	10	9	10	10	9	8
33 to 50	11	5	8	3	3	0	3	5	1	1	2	1	1	4	1	4	1	3
1 to 32	5	0	2	1	0	0	0	1	2	1	1	1	0	0	0	0	0	0
0	0	0	0	12	9	6	1	0	0	8	5	8	3	3	1	2	2	3
	20	20	20	18	14	8	15	19	20	13	12	12	14	16	12	16	12	14
M%	39	60	54	18	20	25	80	79	85	24	41	18	71	67	85	73	76	64

Note.—See Table 1.

* A vs. C: $\chi^2 = 12.12$, $n = 1$, $P > .001$; A vs. N: $\chi^2 = 3.96$, $n = 1$, $P > .05$. All other testable comparisons not significant.

are needed. First, it will be noted that triple coalitions also occur, especially in Type I. Although these might be omitted in determining the incidence of alliances, it facilitated analysis to prorate them among the participants. Second, since the method of drawing weights did not guarantee equal distribution among the players (except in Type I), each pairing did not necessarily have the same opportunity to occur. For example, the 3-2 combination in Type II could happen between C and N only in 16 of the 20 groups. Furthermore, within a triad, a given player might draw a particular weight one, two, or three times. The data,

TABLE 3

INITIATIONS AND RECEPTIONS BY THREE MOTIVE CONDITIONS IN RELATION TO OPPORTUNITY WHEN THE "WEAK" POSITION IS DRAWN IN TYPES II AND III

	Initiations						Receptions					
	First Offer			Final Coalition			First Offer			Final Coalition		
% Opportunity	A	C	N	A	C	N	A	C	N	A	C	N
51 or more	9	7	5	9	6	3	4	6	4	3	6	7
33 to 50	1	2	8	1	3	7	4	6	10	6	4	6
1 to 32	2	2	2	1	2	5	1	4[*]	3	6	4	6
0	7	9	5	8	9	5	10	4	3	9	5	4
	19	20	20	19	20	20	19	20	20	19	20	20
M%	46	37	42	44	35	33	26	46	47	26	42	45

	χ^2	n	P	χ^2	n	P	χ^2	n	P	χ^2	n	P
A vs. C	.62	1	$<.50$	1.23	1	$<.30$	4.51	1	$<.05$	2.12	1	$<.20$
A vs. N	2.12	1	$<.20$	4.78	1	$<.05$	6.22	1	$>.01$	3.29	1	$>.05$
C vs. N	.48	1	$<.50$	1.30	1	$<.30$	.44	1	$>.50$	.16	1	$<.70$

[*] Division for C vs. N.

therefore, are cast in terms of the percentage of the opportunities in which a particular pair could form a coalition. Chi square was used in statistical analysis, although means are also presented. The peculiarities of the distributions made *t* tests inappropriate.

Two striking points are evident: (*a*) all pairs are about equally represented, except in Type I, where CN is significantly more frequent than AN; (*b*) as in other experiments in this series, the alliances are formed in accordance with initial strength, the weak pairings occurring in the great preponderance of instances. From this standpoint, motivational differences among the players have very little effect upon play.

Much the same result appears when participation of each motivational category is determined, as shown in Table 2. Except in Type I, where *S*s high in achievement are involved in alliances significantly less often than either of the others, there are no significant differences.

Analysis of the characteristics of play, however, reveals a distinction between *S*s high in achievement and high in nurturance. Table 3 summarizes performance for the "weak" players in Types II and III (i.e., drawing Weight 2 in those games). Each entry in the table is based upon the occurrence of coalitions as percentage of opportunity. "Initiation" refers to proposing that an alliance be formed; "reception" means the member to whom the offer was made. "First offer" signifies the opening gambit, and "final coalition" means the alliance that terminated play in that game. Cases where only one offer was made and accepted are assigned to both classes of events. Scrutiny of Table 3 discloses that *S*s high in achievement tend to be initiators, with the other two acting as recipients.

We may quickly dispose of Type IV. As in other experiments, coalitions are formed much less often than in other types (here, in 30% of the games, all but one of the alliances involving the all-powerful players). It is worth noting that in line with our hypotheses, *S*s high in achievement permitted only 19% of these outcomes when they held the strong weight, but that the controls and *S*s high in nurturance allowed 41% each.

The question of deals actually consummated can be combined with the general matter of winning. Analysis showed no significant distinctions, such as one kind of *S* receiving a

TABLE 4

PERFORMANCE OF THREE MOTIVATIONAL GROUPS FOR TOTAL SERIES OF GAMES (N = 20 Triads)

Index of Accomplishment	Achievement	Control	Nurturance
1. Mean points assigned[a]	392.50	382.50	425.00
2. Mean points gained	393.53	389.86	416.13
3. Mean gain or loss over assigned points	1.03	7.36	−8.87
4. Mean rank in triad			
Assigned points	2.2	2.0	1.8
Points gained	1.9	2.1	2.0
Net over assigned	1.9*	1.9	2.2

[a] Points assigned as follows: Type I: 33.3 points per S; Type II: 50 points for each "2" drawn; Type III: 50 points for each "2" or "3" drawn; Type IV: 100 points for each "4" drawn.

* A vs. N: $\chi^2 = 5.58$, $n = 1$, $P > .01$.

greater share of the points (indeed, all alliances in Type I were on a 50–50 basis, and disproportionate deals were too infrequent in other types to permit statistical analysis); nevertheless, *S*s high in achievement appeared consistently to benefit during the course of play. In consequence, several efforts were made to assess the comparative success of the three motive categories over the entire series of 12 games. Table 4 gives some of these indexes. The first three rows depict the situation in terms of points earned. Because of discrepancies in actual potential (power), a gross adjustment was made by ascertaining what the outcome would be if each *S* played strictly according to initial strength (as explained in the note to the table). The second row shows the actual total accumulated, were outcome in the total series of games to be summed. The third row gives the average gain or loss when points won are subtracted from points assigned. Although none of the differences is significant, nevertheless, it can be seen that despite the lower potential of *S*s high in achievement, compared to those high in nurturance, the former end up quite well.

Finally, the fourth comparison concerns intratriad relationships. It is clear that *S*s high in achievement gain enough points somehow to move, on the average, from the lowest potential rank to the highest actual rank at the end of the series. It is striking that the differences in average rank are not statistically significant, but that when *S*s are compared in terms of rank in net gain over assigned points, those high in achievement are significantly different from their rivals who are high in nurturance ($\chi^2 = 5.58$, $P < .01$). In fact, *S*s high in achievement ranked first, in these terms, in 10 of the 20 triads, compared to 3 for *S*s high in nurturance. The strategy whereby this was accomplished will occupy us in the next section.

DISCUSSION

Two major points emerge from the results of this experiment: first, that perceived difference in power is a very compelling determinant of behavior, outweighing motivational differences and, second, that even despite this factor, persons high in achievement manage to improve their positions relative to the other participants.

The first point is fully in line with other experiments, since we have yet to discover a condition that overrides the outcome predictable on the basis of relative strength. A glance at Table 4 reveals that despite the deeper considerations to be discussed, *S*s high in nurturance actually acquired the most points, as reflected in their having fortuitously occupied the most favorable position in terms of initial strength. Table 1 and 2, also, show how the weak pairings vastly exceed those involving a strong member (Table 1) and how participation in coalitions by individual *S*s corresponds to their initial power (Table 2).

It is remarkable, nevertheless, that motive conditions are related to style of play within this general principle, for *S*s high in achievement are able to arrive at a much better final outcome than their initial potential would indicate. They tend to be initiators of offers to ally when they are weak, whereas their fellow players tend to be recipients of offers (Table 3). In Type I, on the other hand, *S*s high in achievement are significantly less often involved in alliances. As a consequence, motive conditions seem to operate differently when participants are equal in strength, compared to situations where power differences exist. If, however, we were to consider that the course of play itself may bring about some awareness of the other players' characteristics—and we would suppose that this would occur early—

then the ostensible all-equal situation would become modified to the degree that *S*s high in achievement acquire an initial advantage, so that, as shown by Hoffman, Festinger, and Lawrence (1954), the other two would tend to ally against him. (An additional point is the fact that scarcely any bargaining occurred in Type I; it should also be recalled that differences in Type I represent an excess over triple alliances, 47% of the total.) In this sense, Type I may not be an exception to the strategy of playing by initial strength.

Why were *S*s high in achievement relatively successful in capitalizing upon their assets? This problem requires that we focus attention upon the complex of factors within the triad itself. By no index does the player high in achievement actually win the competition (except when he would be predicted to do so on the basis of his actual potential; the *S* with the highest potential won in 16 of the 20 groups). Rather, the person high in achievement succeeds in improving his position relative to the other two. No one of the devices by which he does so attains significance by itself, but all work in the same direction. He tends to seek alliances more often when he is strong as well as weak (too infrequent, however, for statistical analysis) and thus gains more alliances of this type, refuses more often to form alliances when he is all-powerful, is willing to accept a smaller proportion of the prize in return for an alliance, and so on. We have been led to the interpretation that the individual high in achievement seeks to manipulate the internal relationships of the triad to maximize his chances of winning. But no single tactic emerges, except the tendency to be an initiator; rather, strategy appears to be a function of all the conditions present in a particular group. For example, in Type I, he may be content with a 50-50 deal if he can get it; failing this, he may propose a triple alliance, thus gaining at least a third of the points. On the other hand, if he perceives a tendency for the other two to ally against him, he may quickly accept a tentative triple alliance, proposed by another member, without seeking a better deal. Similarly, he may either seek a comparatively small share of the prize when he draws the strong position or he may be satisfied to see what the other two propose; if they do not form an alliance he gains all the points (a rare occurrence, but favoring the player high in achievement). If this interpretation is sound, then the primary consideration for the person strongly motivated to achieve has little to do with the actual score, but much more with his standing relative to his competitors seeking to improve his position in whatever way is feasible under the circumstances. In this fashion, he apparently succeeds in weaving all the possibilities into a relatively more successful style of play than is the case for his opponents. Is this not always typical of the successful player? What baseball player is really concerned about the size of the final score? He knows that a score of 2 to 1 counts just as much as a score of 19 to 0.

SUMMARY AND CONCLUSIONS

An experiment was conducted to determine the effect of motivational differences on play in a competitive three-person game where initial strength of participants varies. In each of 20 groups, one member was high in achievement but low in nurturance, one high in nurturance but low in achievement, and one intermediate in both variables. Each group played 12 games, three each of four types of power pattern. Performance was analyzed in terms of coalitions, participation in coalitions, initiation and reception of offers to ally, and final outcome.

In general, the course of play corresponds much more closely to the pattern of initial strength than to motivational differences. Within this condition however, *S*s high in achievement tend to initiate offers, whereas those high in nurturance tend to receive them. Assessment of relative success at the end of play shows that *S*s high in achievement improve their rank within triads to a significant degree. It is suggested that the person high in achievement employs a strategy, adapted to immediately relevant conditions, that serves to capitalize upon his assets.

REFERENCES

CAPLOW, T. A theory of coalitions in the triad. *Amer. sociol. Rev.*, 1956, **21**, 489-493.

CHAPMAN, L. J., & CAMPBELL, D. T. An attempt to predict the performance of three-man teams from attitude measures. *J. soc. Psychol.*, 1957, **46**, 277-286.

DeCharms, R. Affiliation motivation and productivity in small groups. *J. abnorm. soc. Psychol.*, 1957, **55**, 222–226.

Edwards, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

Hoffman, P. J., Festinger, L., & Lawrence, D. H. Tendencies toward group comparability in competitive bargaining. *Hum. Relat.*, 1954, **7**, 141–159.

Mills, T. M. Power relations in three-person groups. *Amer. sociol. Rev.*, 1953, **18**, 351–357.

Mills, T. M. The coalition pattern in three-person groups. *Amer. sociol. Rev.*, 1954, **19**, 567–667.

Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

Vinacke, W. E. Sex roles in a three-person game, in press.

Vinacke, W. E., & Arkoff, A. An experimental study of coalitions in the triad. *Amer. sociol. Rev.*, 1957, **22**, 406–414.

Wolff, K. H. *The sociology of Georg Simmel.* Glencoe, Ill.: Free Press, 1950.

(Received December 12, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

STATUS AND CONFORMITY TO PRESSURES IN INFORMAL GROUPS[1]

O. J. HARVEY
University of Colorado
AND CONRAD CONSALVI
Vanderbilt University

The primary focus of this study was on the differential reactions of the leader, the second ranking member, and the lowest status member of an informal group to pressures from other members of the group. The reciprocal side of this question, the differential influence of members in these three status positions on the evaluations of the other members, was also investigated together with the relative weight of each in determining the group norm.

A member's striving to gain and maintain status in one of his more important reference groups should be among the key determinants of his conformity to conflicting group judgments. The higher his motivation to attain these ends, and hence the greater his dependency on the group for motive satisfaction, the greater should be his susceptibility to group influence. This possibility assumes the status system of the group to be sufficiently open that a member perceives as real the possibility of either moving up the status hierarchy or of being displaced from his present station by other aspirants to his position. Viewed this way, the central question of this study becomes that of whether or not the three selected status occupants are differentially motivated to improve and maintain their positions in their reference group. These three members were selected as the main figures of the study because there exist plausible reasons for attributing greatest striving and conformity to each of them.

The hypothesis that motivation is a simple function of the magnitude of distance from the goal would favor the possibility of greatest striving and conformity on the part of the lowest status man, in line with results of Fenchel, Monderer, and Hartley (1951) in their test of Benoit-Smullyan's (1944) status equilibration hypothesis. In their study, the reference groups in which subjects displayed the greatest status striving were those in which they perceived their status as lowest. And yet, the very remoteness of the lowest status man from the top position might result in his being less motivated to move up the status ladder and consequently less sensitive to group pressures, especially if his position were a secure one. If, on the other hand, his position is a tenuous one, and if, as Thrasher (1927, p. 332) long ago averred, "... any status in the group is better than none....," then the lowest status man might manifest greatest conformity since any further status loss would jeopardize his membership in the group.

The second ranking person could prove to be the member on whom the goal of the top position exercises greatest motivational pull. Studies relating to goal gradient, level of aspiration, and class mobility seem consistent with this possibility, as is William James' observation on "... the paradox of a man shamed to death because he is only the second pugilist or the second oarsman in the world. That he is able to beat the whole population of the globe minus one is nothing; he has 'pitted' himself to beat that one; and as long as he doesn't do that nothing else counts." (Knight, 1950, pp. 103–104).

The findings of Dittes and Kelley (1956) suggest the greatest conformity for the second status man, if the degree of one's perceived acceptance in a group can be assumed to relate highly to his status in the group. These authors found the highest incidence of conformity to group evaluations to occur among those individuals who had been led to believe that their degree of acceptance by the other members was average, which in this case was next to the highest of the four degrees of induced acceptance.

---

[1] This study was carried out under support of the Group Psychology Branch, Office of Naval Research, Contract Nonr 2149(02) with Vanderbilt University. Thanks are extended to ONR for its support; to John Whaley for his aid in collecting the data; and to the officials and boys at Tennessee State Vocational Training School, Jordania, for their interest and participation in the study.

The leader's behavior should be less affected by striving for the top position than that of either of the other two status members by virtue of his having attained that goal. To the extent that conformity relates to striving for the goal, his need for conforming behavior should be at a minimum. The attainment of the top position, however, exposes the leader to pressures from a new source, that of maintaining his leadership status. If his position is secure, the leader can perhaps afford to deviate further from the behavior of the other members (Kelley, 1951); but if insecure in his position, the leader might be the most conforming to pressures from the group because of the greater psychological distance he could fall if displaced from the top of the hierarchy.

The studies of Sundby (1959) and Harvey and Rutherford (1958) also bear on this problem. Higher status members were found to be more yielding than those of lower status to majority opinion when the issue involved was relevant to the group (Sundby, 1959). But higher status persons were less influenced by deviant evaluations of other high status members on an issue of minor significance to the group (Harvey & Rutherford, 1958). However, in neither of these studies were leaders' responses separated in the analyses from those of other higher ranking members.

The differential influence of the leader, the second ranking member and the lowest status member in both the formation and the change of group norms was also investigated. The results of studies concerned with the importance of status to source effectiveness, in the informal group (Harvey & Rutherford, 1958) and in larger social units (Hovland, Janis, & Kelley, 1953), led to the simple assumption that the relative influence exerted in both these instances would follow the status positions directly. The design of the study was not ideally suited, however, for answering the latter questions.

METHOD

Subjects

Twenty-seven informal cliques (122 *S*s in all) were selected sociometrically from a state training school for delinquent boys. Nine groups, each with an average of 4-5 members, served in each of three conditions of source influence. Each group served only in one source condition.

Cliques in the training school were selected for study because of the assumption that status within them was very important to the members and that the resulting pressures against the deviant would be pronounced. The cliques were ascertained by a four-item sociometric questionnaire which asked for preferences from among all the boys in the school as teammates and as team captains for a baseball team and for the team in which they preferred to perform the experimental task. Only the responses on the two items having to do with preferred captains or leaders were considered in determining status within the group.

Task and Apparatus

The experimental task was the judgment of the distance between two simultaneous flashes of light in a completely dark room. Two sets of such stimuli were used for each group. The leader or the second or the lowest ranking member (*source*) was exposed at appropriate stages to one set of stimuli while the remaining members of the group (*members*) were being presented with the other pair of light flashes. In all conditions, the distance between the stimulus lights was 12 in. for the members, while for the source the distance between the lights was either 12 or 48 in., depending on the condition. Unknown to the *S*s, a partition separated the two sets of stimulus lights so that the source and members could be exposed to different stimuli while assuming they were seeing the identical stimulus. The use of the partition avoided the need for collaborators, thus leaving *S*s free to interact naturally and to impose the pressures normally applied in their particular group. Special caution was exercised by *E* throughout the experiment to prevent *S*s' bumping into the partition and becoming suspicious of the procedure of the study.

Activation of the lights was controlled by circuits in two Hunter Decade Interval Timers. The two lights, on the same circuit, flashed on and off at the same time. Exposure time was .4 sec. and the time between trials was either 15 or 30 secs., 15 secs. for private and 30 secs. for public judgments.

Procedure

The sociometric questionnaire was administered in the boys' own dormitory approximately two weeks before the selected groups participated in the experiment. The questionnaire was introduced with the explanation that its purpose was to pick groups of friends who would like to participate in a forthcoming experiment. The experiment was described as a study being sponsored by the Navy and concerned with the visual accuracy of individuals working in small groups in complete darkness. Cash prizes were promised for the three groups that made the most accurate judgments on the experimental task. The cash awards promised for the first, second, and third most accurate groups were six, four, and two dollars, respectively, for each person in the group. It was emphasized that the scores would be based upon judgments of the entire group so that no single individual could win a prize without his whole group having won. The offer of cash awards and emphasis upon the group as the winner was intended to increase the likelihood that the boys would take the study seriously, make careful selections in the sociometric choices, and hence be more prone to apply group pressures against persons who made deviant judgments in the subsequent experimental situation.

The conditions according to which the cash prizes would be awarded were re-emphasized to each group as it came to take part in the experiment. Prior to going into the experimental room, each person within a group was assigned a number that represented the order in which judgments of the light distance were to be called out. The source, one of the three selected status occupants, was always assigned the highest number so his judgment would be made last. *S*s were then led into the experimental room, a large room in one of the dormitories, and seated in the reverse order. The source, always seated first, was placed on one side of the partition and the members were seated on the other side of the partition so that neither could see the stimulus lights of the other.

*S*s were then instructed that their task was simply to judge the distance between the two flashes of light. The source and members, after being advised correctly of the 12-ft. distance between them and the lights, were then presented the 12-in. stimulus lights and told the actual distance between them in order to reduce further the intersubject variability in scales of judgment. To prevent the sample presentation from stereotyping the *S*s' responses, *E* pointed out that the pairs of lights that were to follow might all be the same distance apart, or no two pairs might be the same.

The procedure from here on was divided into three phases:

1. This phase was concerned with the *differential contribution* of the highest, middle, and lowest status members to the formation of a group norm. The aim was to have each person first reach an individual standardization in his judgments of the stimulus and then have a group norm formed in which these personal norms would be differentially reflected. To form the personal norm all *S*s were presented the same stimulus, the light flashes of 12-in. distance, for 20 trials with a time interval of 15 secs. between trials. In this stage *S*s recorded their own judgments on small pads without knowledge of any other member's judgment. The group norm was formed by presenting the same stimulus for 10 more trials with only one judgment acceptable to all members allowed for each trial and with an interval of 30 secs. between trials for discussion and joint decision. A score of zero, represented to the *S*s as harmful to their chances of winning a prize, was recorded if a group decision was not reached within the 30-sec. period. Group pressures consequently were applied to the point that a zero score rarely occurred, and when it did it was always only for the first trial.

2. This was the phase critical to the main problem. It was concerned with the extent to which the occupants of the selected status positions would deviate from the judgments of the other members; and if there was deviation, with the differential influence of the three sources on shift in judgments by the other members. This objective was approached in two steps. To provide a control measure for both source and members for the subsequent influence situation, all *S*s once again made 20 private judgments of the presented stimuli, which they recorded without discussion. In this block of trials, however, the source and members were presented different stimuli. The source was exposed to the pair of lights 48 in. apart from one side of the partition, while from the other side the members continued to be presented with the 12-in. stimulus lights. Following the 20 private judgments, the same stimuli were judged for 20 more trials under influence conditions in which, instead of being forced to reach a common group norm, each *S* gave his own judgment aloud after each trial. An interval of 30 secs. between trials allowed enough time for *S*s to give their judgments of what they assumed erroneously to be a common stimulus and to register their reaction of approval or disapproval of the evaluations of the others.

3. This last phase was concerned with the *continuation* of the effects of social influence in a situation free of external pressures, or in current terminology, the correspondence between *public* and *private* conformity. For this purpose, the source and other members rendered 10 private judgments of the same stimuli that they had judged publicly in the preceding phase.

One *E* remained in the experimental room with the *S*s during the experiment, while a second *E*, seated in an adjoining room, controlled the presentation of the stimuli.

RESULTS

Differential Influence of Status in Formation of Group Norm

This result was determined by subtracting the personal norm of the leader and of the second and lowest status members in each group, represented by the mean of his first 20 private judgments of the 12-in. stimulus, from the group norm, represented by the mean of the 20 unanimous judgments of the same stimulus by the entire group. The assumption was that the smaller the difference the greater the weight of the respective source judgments in forming the group norm. Results in Table 1 show the difference to have been

TABLE 1

DIFFERENCE BETWEEN THE PERSONAL NORM OF LEADER, SECOND RANKING, AND LOWEST STATUS MEMBER AND SUBSEQUENT GROUP NORM FOR THE 12-IN. STIMULUS

Status[a]	Group Norm–Source's Personal Norm	Comparison of Differences			
			Diff.	*t*	*P*
Leader	3.67	Leader-Second	1.04	1.28	>.20
Second	4.71	Leader-Lowest	1.57	1.39	>.10
Lowest	5.24	Second-Lowest	.45	.30	>.60

[a] $N = 27$ for each of the three sources, a source of each status from each of the total 27 groups which were combined in this comparison.

TABLE 2

SOURCES' MEAN JUDGMENTS OF THE 48-IN. STIMULUS BEFORE AND DURING MEMBERS' INFLUENCE

Source[a]	Before Influence	During Influence	Diff.	Comparison of Differences			
					Diff.	t	P
Leader	47.70	32.08	15.62	Leader-Second	−23.53	3.21	<.01
Second	56.67	17.52	39.15	Leader-Lowest	−4.02	.58	>.50
Lowest	46.87	27.20	19.68	Lowest-Second	−19.47	22.94	<.001

[a] $N = 9$ for each of the three sources, one source from each of the nine groups in each of the three experimental conditions.

TABLE 3

SOURCES' MEAN JUDGMENTS OF THE 48-INCH STIMULUS DURING AND AFTER MEMBERS' INFLUENCE

Source[a]	During Influence	After Influence	Diff.	Comparison of Differences			
					Diff.	t	P
Leader	32.08	40.53	8.45	Leader-Second	1.11	.14	>.50
Second	17.52	24.86	7.34	Leader-Lowest	−1.23	.16	>.50
Lowest	27.20	36.88	9.68	Lowest-Second	2.34	.37	>.50

[a] $N = 9$ for each of the three sources, one source from each of the nine groups in each of the three experimental conditions.

smaller, but not significantly so, the higher the status of the source.[2] Thus, while the data suggest that contribution to norm formation may be a positive function of status, the lack of statistical significance does not permit the difference in subsequent conformity to group pressures to be attributed to the differential weight of different status members in the initial formation of the norm, as some have supposed.

Differential Conformity of the Three Sources

Conformity of sources' to members' judgments was measured by the difference between the sources' mean judgments of the 48-in. stimulus before and during the phase that public judgments of the 12-in. lights were being made by the members. Results in Table 2 show the second ranking source to have been influenced significantly more by members' judgments than was either the leader or lowest status member. The leader was least influenced, but not significantly less than the lowest status source.

Conformity of Sources in the Private Situation Following Members' Influence

It can be inferred from Table 3 that the influence exerted by the members on sources' evaluations continued even when the external pressures were removed. The three sources tended to shift their judgments of the 48-in. stimulus toward the veridical when judging it in private, but none did so significantly nor to an extent significantly different from the others.

Differential Effects of Sources on Members' Judgments

The fact that the three sources differed in the extent to which their judgments deviated from the members' necessitated an analysis of member change in terms of proportions rather than absolute magnitudes: the ratio of member change to the amount of change advocated by the source. The change in each member's judgment of the 12-in. stimulus from the second private to the immediately subsequent influence condition was divided by the difference between the mean of that member's judgment of the 12-in. stimulus in the second private condition and the mean of the source's judgment of the 48-in. stimulus during influence.

In Table 4 is presented a comparison of the relative effectiveness of the somewhat discrepant evaluations of the three sources in changing members' judgments. The higher the status of the source the greater his proportionate influence on members' judgments, although the only significant difference in source effectiveness was between the leader and lowest ranking member ($t = 2.01$; $P < .05$).

Latent Source Effects Following Influence

In Table 5 are presented data concerning changes in members' judgments of the 12-in. stimulus in the third private condition, the condition that followed the members' experience of the more discrepant evaluations by the source. The leader's influence on members' judgments increased when the members were

---

[2] All t tests reported in this study are based on the formula for correlated samples (Wert, Neidt, & Ahmann, 1954, p. 141).

TABLE 4

EFFECTS OF DISCREPANT SOURCE EVALUATIONS ON CHANGING MEMBERS' JUDGMENTS OF THE 12-IN. STIMULUS

Source Condition	Members N	Mean Source Discrepancy	Mean Member Change	Member Change: Source Deviation
Leader	31	10.20	1.53	.15
Second	34	4.65	.35	.08
Lowest	30	6.63	−3.50	−.53

TABLE 5

LATENT EFFECTS OF SOURCES ON MEMBERS' PRIVATE JUDGMENTS FOLLOWING PUBLIC INFLUENCE

Status of Source	Members N	Mean Change		Mean Diff.	t	P
Leader	31	1.27	Leader-Second	1.51	2.01	<.05
Second	34	−.24	Leader-Lowest	1.72	2.12	<.05
Lowest	30	−.45	Lowest-Second	−.21	.38	>.50

removed from the restraints of the public condition, whereas the influence of the other two sources was less in the private than in the preceding public condition. The second-order difference between measures of differential leader influence on members' judgments under conditions of private vs. public judgment, on the one hand, and corresponding measures of differential influence by the second and lowest ranking sources, on the other, was statistically significant, but the difference between the latent effects of the latter sources was not significant.

DISCUSSION

The most important of the foregoing results is probably the demonstration of greatest conformity to group pressures on the part of the second ranking persons, those only one step removed from the leadership position. This finding is compatible with goal gradient theory and the concept of relative deprivation, both of which underlie certain phenomena of level of aspiration, including greater status strivings and conformity of members of the middle class.

The failure of the study to show any clear-cut differences in the influence of the highest, second, and lowest status members in the formation and change of group judgments could well be due to the small size of the sample and to limitations of the experimental design inherent in the focus of the study. An adequate test of the effects of different sources would require their judgments to be made at a controlled and constant distance from the judgments of the other members, a requirement not met in this study because of the differential effects of the group on the sources.

The tendency for the three sources to continue to give decreased judgments after removal of external group pressures suggests that the changes in sources' judgment were due to something other than just public yielding to group pressure, a point similarly illustrated in Sherif's (1935) early experiment. Such change might be unwitting, or the interesting possibility exists that, even though a person may have knowingly modified his judgments to agree with another's, pride may not allow him to switch back immediately to his initial evaluation even when he is free to do so.

An unpredicted but not surprising finding was the tendency for the effects of leaders on members' judgments to increase when the latter were rendered subsequently in private conditions in which the restraints of public surveillance were lessened.

SUMMARY

This study was concerned primarily with the differential conformity of the leader, the second ranking member and the lowest ranking member to pressures in the informal group. The informal groups were 27 cliques sociometrically selected from a training school for delinquent boys.

Under the incentive of a financial reward for the group most accurate in the experimental task, members of each group judged the distance between two simultaneous flashes of light in a dark room. While the other members were judging the distance between two flashes of light 12 inches apart, the leader (or second ranking or lowest status members depending on the experimental condition) estimated the distance between two flashes of light 48 inches apart, which unknown to the *S*s were separated from the other lights by a partition. In this condition, under the assumption all were seeing the same lights, *S*s made their judgments aloud and brought verbal pressures against

the person who rendered the deviant judgments.

The main finding of the study was that the second status member, the one only one step from the top, was significantly more conforming to judgments of the other members than was either the leader or lowest status man. The leader was least conforming, but not reliably less than the lowest status person.

REFERENCES

BENOIT-SMULLYAN, E. Status, status types and status inter-relations. *Amer. socio. Rev.*, 1944, **9**, 151–161.

DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100–107.

FENCHEL, G. H., MONDERER, J. H., & HARTLEY, E. L. Subjective status and the equilibration hypothesis. *J. abnorm. soc. Psychol.*, 1951, **46**, 476–479.

HARVEY, O. J., & RUTHERFORD, JEANNE. Relationship of status in the informal group to influence and influencibility. *Child Develpm.*, in press.

HOVLAND, C. L., JANIS, I., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.

KELLEY, H. H. Communication in experimentally created hierarchies. *Hum. Relat.*, 1951, **4**, 39–56.

KNIGHT, M. (Ed.) *William James.* London: Pelican Books, 1950.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.

SUNDBY, E. Unpublished doctoral study, Vanderbilt University, 1959.

THRASHER, F. M. *The gang.* Chicago: Univer. Chicago Press, 1927.

WERT, J. E., NEIDT, C. O., & AHMANN, *Statistical methods in educational and psychological research.* New York: Appleton-Century-Crofts, 1954.

(Received February 16, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

THE STABILITY OF INTERPERSONAL JUDGMENTS IN INDEPENDENT SITUATIONS[1]

EDGAR F. BORGATTA
Cornell University

WHILE there are many problems involved in the use of ratings and rankings, their pervasiveness in social and psychological research attests to their importance. Thus, rather than to shy away from dealing with ratings and rankings, the need is for more systematic exploration and examination of their limitations. The study reported here brings together into a single experimental design several aspects that have generally received attention separately. The generic form of such a multitrait-multimethod design is discussed by Campbell and Fiske (1959).

The research variables are stratified in three ways:

1. *Multitrait.* The research includes three dimensions of trait description. The scores representing these dimensions are based on prior work indicating that only a limited number of dimensions will account for most of the variance of interpersonal judgments.
2. *Multimethod.* It provides data on interpersonal judgments made in two independent situations.
3. *Multimethod.* It examines the interrelationship of self-ratings, self-rankings, and peer rankings.

The Dimensions of Interpersonal Judgments: Prior Work

An important emphasis in the use of sociometric ratings has been on the content of the criterion on which group members are asked to choose (or rate or rank) each other. Experience in use and direct research have led to a distinction between work (sociogroup) and the friendship (psyche-group) choices as relatively independent; that is, recognition has been given to the fact that the person who is chosen as the leader for a task situation might not be the same as the one with whom social activities might best be realized. Jennings (1947, 1950), for example, gave explicit attention to this analysis. Bales and Slater (1955) and Slater (1955) emphasized a similar distinction in the study of interaction process, under the rubrics of task and social-emotional specialists. In responses gathered after discussion meetings, however, organizing concepts included group member rankings on (*a*) contributing best ideas and (*b*) guiding the discussion, and ratings on (*c*) personal liking for the group members.

While there have been many other classifications of roles and of types of behavior, the above are particularly worthy of note since they stem from large bodies of empirical research. However, while the sets of concepts arising from both sociometry and interaction process analysis reflect an accumulated body of experience, the delineation of dimensions of ratings and rankings appears as a by-product rather than as a consequence of a direct inquiry such as Carter's (1954). On the basis of factor analyses of his data, as well as a number of independently reported studies, Carter suggested that three categories account for most of the variance of individuals as seen in small group interaction: (*a*) individual prominence and achievement, (*b*) aiding attainment by the group, and (*c*) sociability. The relationship of these categories to the theoretical developments in sociometry and interaction process analysis is directly visible. Additional work undertaken to examine Carter's proposed three-factor system substantiates it at least in part, but the suggestion arises that four and possibly more factors may be required in order to account for the major portion of the variance of interpersonal judgments, with the dimensions organized around the concepts of individual assertiveness, sociability, intelligence, and emotionality (Borgatta, Cottrell, & Mann, 1958).

---

[1] This study stems from the program of research titled, "The Variables and Conditions of Small Group Interaction" at New York University. Jonathan Robbin, technical associate, was in charge of machine analysis, and the research assistants were John Stimson, Robert Guerrin, Gerald Marwell, and Ardyth Stimson. This research was supported in whole or in part by the United States Air Force under Contract AF49(638)-195 monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

We may at least conclude that propositions about the stability of judgments must be made with attention to the dimensions of content involved. The importance of this requirement has been particularly emphasized by Cronbach (1955) in his analysis of discrepancy measures.

The Consistency of Peer Judgments: Prior Work

In the utilization of peer judgments, the question of the reliability of scores is a prominent and complex one. For example, if a person is said to have a high score on judgments made by peers on a sociability criterion, the question of reliability is not merely whether scores taken again in the same way for the same group will be relatively stable. Implicitly, the expectation is that the score identifies an attribute of the individual with some degree of generality. Still, it is quite another question to ask about the stability of such scores when they are taken in different groups or situations.

Referring particularly to sociometric measures, Mouton, Blake, and Fruchter (1955) report only two studies of the consistency of judgments under conditions of experimental variation. The first is the study by Bell and French (1950) that employed a single criterion (leadership), in which it was found that with a sample of 25 subjects (*S*s), organized as five-person groups in six successive 35-min. discussion sessions in which no two *S*s participated together more than once, the ranking scores of individuals were for any one session correlated on the average .75 with the average of the other sessions. The second study reported by Mouton, Blake, and Fruchter was their own, with a sample of 33 *S*s organized in three-person groups in two successive 15-min. discussion sessions, but in this case using 12 items or criteria. They concluded from their study that judgments of behavior can be made reliably under conditions of short-term interaction and indicated that the items most reliably judged in their study were leadership, contribution to group decision, and dominance position within the group.

Both these studies emphasize the consistency with which peers characterize the individual, particularly in terms of short-term contact, but they do not focus on the dimensions of content in judgments as such. In the second study just mentioned, the three items reported to be most reliably judged might be expected to correlate at a very high level and thus to represent only one dimension. Other studies that have touched on the consistency of judgments with varying composition of groups, such as that of Bass and Wuster (1953), have reported positive findings but also have not been concerned with examining the problem of dimensions of judgment.

Self and Peer Judgments

In examining consistency over independent situations, there is as yet no systematic report on the relationship of self judgments with those of peers. The data reported here permit the self and peer judgments to be examined in two independent discussion groups, with parallel analyses for three dimensions of judgment. Although the accuracy with which a person perceives the judgments of others of course enters these data indirectly, the data pertain to the correlation of the self-judgment with peer consensus. This is a measure of correspondence rather than of accuracy as such.

In addition, prior self-ratings representing an abstract frame of reference are compared with the judgments made in terms of the concrete sets of other persons in the two independent discussion situations. Thus, this design permits examination of the consistency of self-ratings made "alone" with self-rankings made in the context of specific others, as well as the comparison of self-ratings with peer rankings.

DESIGN OF THE ANALYSIS

The sample employed for this analysis consisted of 99 male *S*s, college sophomores and juniors. Data were collected in three stages with from one to three months between each stage.

The first type of data collected was a self-rating check list of descriptive behavioral characteristics responded to as part of a general orientation test period for *S*s. The check list represents 40 trait descriptions that have been developed and used in extensive research on the rating and ranking of behavior in interaction situations (Borgatta, Cottrell, & Mann, 1958). The check list form asked the *S* to circle a number from 0 to 9 associated with a continuum running from "Definitely does not describe me" to "Definitely does describe me."

The second source of data was from a postmeeting questionnaire, based on the same 40 descriptive behavioral characteristics, that was completed after participation in a five-person discussion. *S*s who were not acquainted participated in the interaction situation for 80 min., with four discussion topics. These topics were selected from an extensive attitude questionnaire called the General Orientations Profile that was administered previously at the orientation test period. Topics were chosen to represent orthogonal attitudinal content in the subject population. Thus, the successive alignments of agreements and disagreements on the topics occurred by chance within a session, and the opportunity was maximized for a person to find a topic in which he had some interest and on which he took some position. In this sense, also, the discussion situation was developed to maximize the possibilities for exposure to discussion participation on a variety of issues in a relatively short period of time. The questionnaire for ranking the group members was administered after the discussion. This provided self-rankings of the *S*s on the descriptive behavioral characteristics and also the peer rankings (minus self-rankings) on the same characteristics.

The third source of data was similar to the second, but was carried out in a somewhat shorter discussion session (three new topics, 20 min. each, with the same orthogonal content as three of the previous topics), in three-person rather than five-person groups, and again in groups composed of persons who were not acquainted. Self and peer rankings on the characteristics were available from this source.

The three scores based on the ratings or rankings employed in this analysis are *Assertiveness*, *Sociability*, and *Emotionality*. Each score is composed of several highly intercorrelated variables found to be consistently associated with orthogonal clusters in factor analysis. The score in each case is the sum of the self-ratings, self-rankings, or peer rankings on the relevant items. The scores are composed of responses to the following items: Assertiveness (A)—Is very active, Initiative, Makes many suggestions, and Assertiveness; Sociability (S)—Shows solidarity and friendliness, Makes others feel he understands them, and Likeability; Emotionality (E)—Is very tense, Gets upset very easily, Tends to be very nervous, Emotionality. The remaining 29 items tend to be distributed on these factors or are involved in less well-defined factors.

RESULTS

Dependencies between the Contents

The matrix of product-moment intercorrelations for all scores in this analysis is presented in Table 1. Since the research focus requires comparisons of individuals in independent situations, the correlations were computed over all 99 *S*s. It would be possible, also, to make more limited examinations of the relationships among measures by computing, say, for each five-person group the intercorrelation matrix of the self-ratings, the five-person self-rankings and peer rankings, and the parallel scores each member received in the independent three-person groups. If this procedure were carried out it would be possible to describe directly the distribution characteristics of the observed relationships over the five-person groups. However, the procedure used appears appropriate in this context since the use of rankings holds constant the sum (or mean) of all ranks within the group, and the variability of group means of self-rankings and peer rankings is restricted within this dependency. Such variability would tend to attenuate relationships between the classes for the pooled data. Consistency of frame of reference between groups as well as within groups is necessary for relationships to be manifest in the pooled data.

A slight but significant relationship occurs among the three self-rating scores with Assertiveness and Sociability positively correlated ($r_{A_1S_1} = .23$) and both of these negatively correlated with Emotionality ($r_{A_1E_1} = -.21$; $r_{S_1E_1} = -.22$). In the self-rankings, both for the five and the three-person groups, Assertiveness and Sociability are positively correlated ($r_{A_2S_2} = .37$; $r_{A_4S_4} = .10$), although for the three-person group the correlation (.10) does not achieve the significance level. In the self-rankings, neither for the five- nor the three-person groups was the relationship between Assertiveness and Emotionality negative ($r_{A_2E_2} = .17$; $r_{A_4E_4} = .05$). On the other hand, the correlation of Sociability with Emotionality is negative and significant for both the five-person ($r_{S_2E_2} = -.20$) and the three-person groups ($r_{S_4E_4} = -.34$).

For the peer ranking scores in the five- and three-person groups, Assertiveness and Sociability are again positively correlated ($r_{A_3S_3} = .51$; $r_{A_5S_5} = .32$), and Sociability is negatively correlated with Emotionality ($r_{S_3E_3} = -.39$; $r_{S_5E_5} = -.41$). The relationship between Assertiveness and Emotionality is indicated to be positive, although slight and significant only for the three-person groups ($r_{A_3E_3} = .09$; $r_{A_5E_5} = .22$).

Although the measures for the three dimensions of content in the ratings and rankings were drawn from previous studies that indi-

TABLE 1
INTERCORRELATION MATRIX
$N = 99$

	Self Ratings			Five-Person Groups						Three-Person Groups					
				Self-Ranking			Peer Ranking			Self-Ranking			Peer Ranking		
	A_1	S_1	E_1	A_2	S_2	E_2	A_3	S_3	E_3	A_4	S_4	E_4	A_5	S_5	E_5
A_1	—	*23*	*−21*	*43*	*31*	−06	*31*	*25*	−12	*22*	14	05	*23*	*31*	−03
S_1		—	*−22*	11	*53*	*−21*	05	07	−19	02	*42*	*−27*	−01	08	−02
E_1			—	−19	*−21*	*29*	10	−08	*22*	−02	*−25*	*26*	−06	−07	17
A_2				—	*37*	17	*55*	*33*	13	*42*	03	07	*39*	10	11
S_2					—	*−20*	05	*20*	*−27*	11	*51*	−16	15	*23*	*−21*
E_2						—	*21*	−08	*43*	14	−17	*50*	*21*	−05	*22*
A_3							—	*51*	09	18	−07	08	*38*	14	05
S_3								—	*−39*	02	04	−05	*23*	*44*	*−23*
E_3									—	05	−17	15	05	*−32*	*44*
A_4										—	10	05	*61*	*22*	*28*
S_4											—	*−34*	00	*27*	−14
E_4												—	16	−08	15
A_5													—	*32*	*22*
S_5														—	*−41*
E_5															

Note.—Decimal points omitted. A, S, and E are the scores on dimensions of Assertiveness, Sociability, and Emotionality. Value of $| r | \geqq .20$ would satisfy a symmetric test of significance at the .05 level with 95 *df*. Significant values are italicized.

cated a high degree of independence among them, in this analysis the three contents tend to be related. The relationship probably depends in part on the fact that the groups had participated together only for a relatively short period of time and the opportunity to communicate had been somewhat restricted. In any set of peer judgments, as the group members get to know one another they should be better able to identify the characteristics of each person and to distinguish accurately between the characteristics in their rankings. When one knows an individual well enough, that is, one can *take into account* his other qualities in ranking him on a specific quality. Consider the following question, for example: "How high is an individual on Sociability?" Taking into account individual differences in activity levels appears to be implicit in the question.

Interrelationships of Self-Judgments Across Situations

The existence of some dependency among the three dimensions of content does not prohibit examining relationships across situations and types of measures. Thus, to begin with, the self-ratings made in terms of abstract reference points are predictive of the ranks *S*s assign to themselves in the five- and three-person groups. The self-rating score on Assertiveness is correlated with the parallel self-ranking score in both discussion groups ($r_{A_1A_2} = .43$; $r_{A_1A_4} = .22$); the self-rating score on Sociability is correlated with the parallel self-ranking scores ($r_{S_1S_2} = .53$; $r_{S_1S_4} = .42$); and the self-rating score on Emotionality is correlated with the parallel self-ranking scores ($r_{E_1E_2} = .29$; $r_{E_1E_4} = .26$). All these values are statistically significant and constitute the largest values, with only minor exception, in the vectors in the intersection matrices. The exception occurs for the correlation between the self-rating score on Sociability and the self-ranking score in the three-person group on Emotionality ($r_{S_1E_4} = -.27$). The negative relationships between Sociability and Emotionality from self-rating scores to self-ranking scores are statistically significant ($r_{S_1E_2} = -.21$; $r_{S_1E_4} = -.27$; $r_{E_1S_2}$ $-.21$; $r_{E_1S_4} = -.25$). The additional significant relationship in these comparisons is a positive correlation of the self-rating score on Assertiveness with the self-ranking score on Sociability for the five-person group ($r_{A_1S_2} = .31$).

Stability of self-ranking scores may also be examined between the five- and the three-person groups. *S*s participated in these groups with different persons, and the restriction of comparison obviously differs when an *S* utilizes four other persons versus two other persons as the frame of reference in identifying characteristics. It is seen, however, that the self-ranking scores in the five-person group correlate with the self-ranking scores in the three-person groups quite substantially, in spite of the obvious restriction on variance involved in the limited number of comparisons in the groups ($r_{A_2A_4} = .42$; $r_{S_2S_4} = .51$; $r_{E_2E_4} = .50$).

The stability of self-ranking scores from situation to situation recalls the problem of contamination of measures that makes unraveling data of an interpersonal type quite challenging. Do the relatively stable indicators mean that a person has a well-defined and stable perception of his role among others or does it indicate, rather, that his place among others is indeed relatively stable and that he is able to identify what it is with some degree of accuracy? This question can be answered in part by examining the self-judgments in terms of the peer judgments, to which we will turn after considering the interrelationship of the peer rankings.

Interrelationship of Peer Rankings Across Situations

As we have already noted, the interrelationship among the dimensions of content used in the peer judgments borders on the substantial in both the five- and the three-person groups. However, when one examines the predictability of peer judgments from five- to three-person groups, one notes that the largest correlations in the intersection matrix occur in the main diagonal and each dimension best predicts itself. Thus, across the situations the correlations are for Assertiveness, $r_{A_3A_5} = .38$; for Sociability, $r_{S_3S_5} = .44$; and for Emotionality, $r_{E_3E_5} = .44$. Peer ranking scores on Sociability are significantly negatively correlated with peer ranking scores on Emotionality ($r_{S_2E_3} = -.23$; $r_{S_3E_2} = -.32$), but the magnitudes are below those of the main diagonal in the intersection matrix. An additional significant relationship occurs in this intersection matrix between the peer rankings on Sociability for the five-person group and the peer rankings on Assertiveness for the three-person group, although the relationship is small ($r_{S_2A_5} = .23$).

We may now examine the question of the predictability of peer ranking scores on the basis of self-rating and self-ranking scores. The peer scores, it should be recalled, do not involve self-rankings.

Self-Judgments as Predictors of Peer Judgments

Self-rating scores on Assertiveness are significantly correlated with corresponding peer ranking scores in the five- and the three-person groups ($r_{A_1A_3} = .31$; $r_{A_1A_5} = .23$). In both cases self-rating scores on Assertiveness are also significantly correlated with peer ranking scores on Sociability ($r_{A_1S_3} = .25$; $r_{A_1S_5} = .31$). Self-rating scores on Sociability, on the other hand, are not significantly associated with the peer ranking scores on Sociability ($r_{S_1S_3} = .07$; $r_{S_1S_5} = .08$). For the five-person groups, self-rating scores on Emotionality are significantly correlated with peer ranking scores on Emotionality ($r_{E_1E_3} = .22$), and the relationship is maintained in the three-person group ($r_{E_1E_5} = .17$) although it falls short of significance.

One notes first that self-judgments employing a specific group of individuals as a frame of reference lead to more substantial predictive correlations for the peer ranking scores in the groups than do self-ratings made in the abstract. Thus, for the five-person groups, self-ranking scores on Assertiveness are substantially correlated with peer ranking scores ($r_{A_2A_3} = .55$). Interestingly, these self-ranking scores in the five-person groups also are significantly correlated with peer ranking scores for the three-person groups, and again at a relatively substantial level ($r_{A_2A_5} = .39$).

The findings for self-rankings in the five-person groups in terms of Sociability and Emotionality parallel those of Assertiveness; that is, self-rankings on Sociability are significantly correlated with peer rankings in the five- and three-person groups ($r_{S_2S_3} = .20$; $r_{S_2S_4} = .23$), although for Sociability the magnitudes are smaller. Similarly, self-ranking scores on

Emotionality for the five-person groups are correlated with peer ranking scores on Emotionality ($r_{E_2E_3} = .43$; $r_{E_2E_5} = .22$).

There are additional correlates as well: self-ranking scores on Assertiveness for the five-person groups are correlated with peer ranking scores on Sociability in the five-person groups ($r_{A_2S_3} = .33$); the parallel relationship occurs for three-person groups ($r_{A_4S_5} = .22$) but does not extend between self-ranking scores of the five-person groups and peer ranking scores for the three-person groups ($r_{A_2S_5} = .10$), or vice versa ($r_{A_3S_4} = -.07$). For the five-person groups, self-ranking scores on Emotionality are significantly correlated with peer ranking scores on Assertiveness ($r_{E_2A_3} = .21$), and also to peer ranking scores on Assertiveness in the three-person groups ($r_{E_2A_5} = .21$). However, the parallel relationships for self-ranking scores on Assertiveness for the three-person groups are not significant ($r_{E_4A_3} = .08$; $r_{E_4A_5} = .16$). Self-ranking scores on Sociability for the five-person groups are significantly negatively correlated with peer ranking scores on Emotionality for both the five- and the three-person groups ($r_{S_2E_3} = -.27$; $r_{S_2E_5} = -.21$); the parallel relationships for the three-person groups are not significant ($r_{S_4E_3} = -.17$; $r_{S_4E_5} = -.14$).

Examining now the self-rankings in the three-person groups, self-ranking scores on Assertiveness are positively correlated at a substantial level with peer ranking scores for the three-person groups ($r_{A_4A_5} = .61$) and, also, although at a much lower level, with the Sociability and the Emotionality peer ranking scores ($r_{A_4S_5} = .22$; $r_{A_4E_5} = .28$); self-ranking scores on Sociability are positively correlated with peer ranking scores on Sociability for the three-person groups ($r_{S_4S_5} = .27$); while the correlation between self and peer ranking scores on Emotionality is in the correct direction, the magnitude is not sufficient to satisfy the criterion of statistical significance ($r_{E_4E_5} = .15$).

Finally, self-rankings for the three-person groups do not appear to be efficient predictors of peer judgments in the five-person groups. It should be noted, however, in this intersection matrix all values in the main diagonal are positive although none are of sufficient magnitude to satisfy the significance criterion ($r_{A_4A_3} = .18$; $r_{S_4S_3} = .04$; $r_{E_4E_3} = .15$).

DISCUSSION AND SUMMARY

Self-rating scores appear to be direct predictors of the self-ranking scores that are made after five- and three-person discussion situations (median $r = .35$), independently of the dimensions of content. This finding could result simply from an individual's self-image or bias that he carries over into his self-perception when ranking himself in regard to any set of specific others. That is, if he judges himself to be relatively high in an abstract sense (self-rating), he may assign himself a rank that is high without attention to the qualities of others, whether he does this unconsciously or deliberately. This is not, it should be noted, the elevation problem that Cronbach (1955) discusses, i.e., the tendency to rate all persons high, low, or at some given level, since the mean rank for each set of judgments is by definition the same.

The conception of an abstract self-image or bias in self-perception is not easily detached from the question of stability of qualities and accuracy of perception. If persons are stable in their qualities, the stability in independent situations would be dependent upon the equivalence of sets of persons in the different situations. Thus, the variation of the characteristics of sets of group members must be thought of as an important source of variance in the "unreliability" of judgments in independent situations. But beyond this consideration, in order for there to be reliability there must be accurate perception of the qualities of the group members in the independent situations. The prediction diagonal of peer judgments from five- to three-person groups (A_3A_5, S_3S_5, E_3E_5) tends to be positive and nontrivial (median $r = .44$), thus establishing the relative stability of ranks in independent groups, irrespective of the dimensions of content. Dependently, this also means that group members must view themselves consistently and with a common frame of reference. Since the prediction diagonals from self-rating and self-ranking scores to peer ranking scores (A_1A_3, S_1S_3, E_1E_3; A_1A_5, S_1S_5, E_1E_5; A_2A_3, S_2S_3, E_2E_3; A_2A_5, S_2S_5, E_2E_5; A_4A_3, S_4S_3, E_4E_3; and A_4A_5, S_4S_5, E_4E_5) tend to be positive and nontrivial (median $r = .22$), group members must also have accurate perception of their qualities according to the group consensus.

A point that needs to be emphasized here is that the respondents were not asked to predict how others would rate or rank them. Thus, except for logical dependency indicated, the ability of the person to predict judgments others make is not involved. This was not the task assigned.

As for the magnitudes of predictions, it appears that correlations of self-rating with self and peer rankings are larger for the five-person groups (A_1A_2, S_1S_2, E_1E_2; median $r = .43$, A_1A_3, S_1S_3, E_1E_3; median $r = .22$) than the three-person groups (A_1A_4, S_1S_4, E_1E_4; median $r = .26$; A_1A_5, S_1S_5, E_1E_5; median $r = .17$). This result accords with the view that the variation of sets of group members restricts the reliability of judgments, i.e., a more general base of judgment is available in five- than in three-person groups since sampling expectations dictate less variation in the means and variances of samples of larger size. In addition, of course, the concrete frame of reference for judgment is larger.

Since dimensions of content were employed that tend to be defined orthogonally in factor analytic studies, the generality of findings could be examined over these dimensions. In general, findings tended to occur in parallel, supporting the notion that principles may be observed without regard for the content but also emphasizing the need for specific and inclusive dimensions in describing characteristics of persons.

REFERENCES

BALES, R. F., & SLATER, P. E. Role differentiation in small decision-making groups. In T. Parsons, R. F. Bales, et al. *Family, socialization and interaction process.* Glencoe, Ill.: Free press, 1955.

BASS, B. M., & WUSTER, C. R. Effects of the nature of the problem on LGD performance. *J. appl. Psychol.*, 1953, **37**, 96–99.

BELL, G. B., & FRENCH, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 764–767.

BORGATTA, E. F., COTTRELL, L. S., JR., & MANN, J. H. The spectrum of individual interaction characteristics: An interdimensional analysis. *Psychol. Rep., Monogr. Suppl.* **4**, 1958, **4**, 279–319.

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81–105.

CARTER, L. F. Recording and evaluating the performance of individuals as members of small groups. *Personnel Psychol.*, 1954, **7**, 477–484.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

JENNINGS, HELEN H. Sociometric differentiation of the psychegroup and the sociogroup. *Sociometry*, 1947, **10**, 71–79.

JENNINGS, HELEN H. *Leadership and isolation.* (2nd ed.) New York: Longmans, Green, 1950.

MOUTON, JANE S., BLAKE, R. R., & FRUCHTER, B. The reliability of sociometric measures. *Sociometry*, 1955, **18**, 7–48.

SLATER, P. E. Role differentiation in small groups. *Amer. sociol. Rev.*, 1955, **20**, 300–310.

(Received February 27, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

STIMULI RELATED TO STRESS, NEUROTICISM, EXCITATION, AND ANXIETY RESPONSE PATTERNS:

ILLUSTRATING A NEW MULTIVARIATE EXPERIMENTAL DESIGN[1]

RAYMOND B. CATTELL AND IVAN H. SCHEIER
University of Illinois

THIS research aims at three targets:

1. To confirm and extend our knowledge of the various responses to what may broadly be called "threat," and which include, according to present evidence, at least the three distinguishable states of anxiety, stress response, and neurotic response.

2. To see if lawful relations can be found between the magnitude of response along these defined factor dimensions and the nature and magnitude of certain commonly occurring "stressful" or "threatening" situational stimuli.

3. To illustrate and gain experience of the new *condition-response design* in multivariate experiment proposed by Cattell (1952) for combining the advantages of factor analytic and controlled experiment. The example here involves the measurement of 86 young men on 69 manifestations of anxiety, stress, and neurotic response on two occasions, differing by presence or absence of (*a*) imminence of an important examination, (*b*) a group "stress interview" by a questionnaire on emotional problems, and (*c*) the challenge of a punishing physical ordeal.

A survey (Cattell & Scheier, in press-a) of recurrent experiments in the area commonly referred to as that of "anxiety" or "stress" response indicates certain persistent obscurities: (*a*) confusion as to whether stress, anxiety, etc., are to be defined by the nature of the stimuli or of the responses, (*b*) in either case, the use of arbitrary and conflicting choices of variables to measure or manipulate experimentally, and (*c*) undue dependence upon a single test or variable "to represent operationally" anxiety, stress, etc., when a priori it is likely that the variance of any such single variable will be composite, and the measure of anxiety or stress requires representation in a single battery of several variables demonstrated to occur in the response pattern.

The methodologically inadequate dependence on a *univariate* approach is illustrated by the work of Taylor (1951) and a succession of similar studies (Hilgard, Jones, & Kaplan, 1951; Ramond, 1953; Spence & Beecroft, 1954; Spence & Farber, 1953; Taylor & Spence, 1952), comparing learning or conditioning for high and low anxiety groups, established as such on the basis of scores on one questionnaire said to be a face-valid measure of anxiety (Taylor, 1953). Almost all such studies employ only one or two tests alleged to represent anxiety, and then seek a relationship between this "anxiety" and one or two other variables, such as conditioning or environmental stress, presumed to have some connection with anxiety.[2] Unfortunately, the correlations among such confidently employed indicators of anxiety are often unknown, or when known, turn out to be low.

By contrast, the multivariate approach first seeks to discover and define what general factor or factors exist, as empirical constructs, among a wide range of semantically and theoretically alleged psychological (and physiological) manifestations of anxiety. Regardless of how the discovered factor is named, a battery of tests for such a factor is then known to measure a unitary response pattern. Researches to locate such unitary factors in the area of stress and anxiety were completed prior to the present research, and they have replicated beyond doubt the pattern of (*a*) a general Anxiety factor, symbolized in the universal index of factors as U.I. 24 (Cattell, 1957a, 1957b), (*b*) a stress response factor, symbolized as P.U.I. 4 (Cattell, 1955a, 1957a), and (*c*) a neurotic debility or "Low Neural Reserves" factor, U.I. 23(−) (Cattell, 1957a; Cattell & Scheier, in press-b). Elsewhere (Cattell, 1957a; Cattell & Scheier, 1958a, in press-a, in press-b) and in subsequent sections of this paper, we document our claims to the interpretation of these factor dimensions as constituting the firm, functionally unitary cores in the common concepts of Anxiety, Stress, and Neurotic Debility.

---

[1] This article reports part of a study conducted under contract to the Research and Development Division, Office of the Surgeon General, Department of the Army.

[2] The fact is that many (including, apparently, J. A. Taylor herself [1953]) do assume that the Taylor Scale by design and definition is a measure of "anxiety." A recent tendency to hedge on this position by disowning the claim to "anxiety" and simply calling it "drive strength" or one type of drive strength merely, repeats and accentuates the logical offense. Actually, there is evidence that drive strengths can be separately measured in man (Cattell, 1957a), but the anxiety measure, though connected in a special way therewith, is conceptually and factorially distinguishable.

Despite the satisfactorily definitive outcome of the researches confirming these factors, the main ten factor analytic experiments were only in terms of individual differences between people, at any one time, and of responses to fixed situations. That is to say, the simple structure patterns appeared in regular cross-sectional, *R*-technique analyses, defining *traits* rather than *states*. Since then, however, two *P*-technique researches, factoring serial measures in single individuals, have shown patterns so similar to those found in *R* technique that it is reasonable to call them by the same names and to expect that part, at least, of the individual differences existing at a given moment of cross-section (*R* technique) are due to differences in these temporary states. Thus, an anxious person is partly characterologically so and partly a person temporarily in an anxious state, the two patterns being so similar that they cannot be factorially separated—so far.

However, in both *R*- and *P*-technique studies the situations productive of anxiety have been left to life happenings. In obtaining the true multivariate response patterns, which the univariate experimenter can never hope to see, we have hitherto sacrificed the connection with a stimulus in which he is interested. Such initial attention to the definition *as a response pattern* seems to us the proper strategic sequence, and we now propose to build upon these results by the new *condition-response* multivariate design (Cattell, 1952), which links factor analytic and analysis of variance designs, and promises to connect these response patterns with the stimuli which generate them. It also permits an independent contribution to the definition of the response pattern, for whereas *P technique* takes one person over many occasions, thus sampling well for occasions but poorly for people, the *incremental R technique* (embedded in the total condition-response design) samples only two occasions but performs the usual adequate sampling of people. If these approaches converge on a congruent definition of a single complex response pattern, we can be confident of having something transcending most sampling distortions.

The general theory of condition-response design is given elsewhere (Cattell, 1952), but its nature will be illustrated in the following account of actual experimental procedures. Essentially, it measures each subject under two conditions (which would be high and low either on situational stimulus levels, as here, or internal conditions in, say, a psychophysiological experiment). At the same time, it measures all of a considerable variety of response variable magnitudes and then correlates changes in stimulus magnitudes with changes in response magnitudes, instead of relating significance of difference in one to difference in the other, as in analysis of variance. Upon factoring this matrix, one obtains both (*a*) evidence (or, in our case, confirmation) for what unitary response patterns, e.g., anxiety, stress, etc., may exist in the given range of dependent variables, and (*b*) the significance and magnitude of the relations of these unitary response patterns to the independent variables constituted by the stimulus conditions.

As pointed out above, evidence of the first can be obtained by *R* technique above, or *P* technique. Woodrow (1938) has factor analyzed increment scores in the learning and mental ability domain, and several *R* and *P* studies have recently been carried out in personality data, but to our knowledge the present study is the first to perform incremental *R* technique in personality research in a design with manipulated, randomized, independent variables in what has hitherto been a theoretical proposition for a condition-response design.

DESIGN AND PROCEDURE

S's were 86 male undergraduates: 52 unselected in relation to anxiety, 18 of high anxiety, and 16 of low anxiety.[3] The same 72 variables were scored for each of the 86 *S*s on Occasion 1, and then again on Occasion 2, four weeks later. Sixty-nine of the measures entered in the matrix were *changes* in score from Occasion 1 to Occasion 2 on personality "response" variables; three scores were *changes* in the magnitude of stimuli, representing "massive" environmental "stress" conditions. The change score was always computed[4] as Occasion 2 minus Occasion 1.

Most of the response variables are well known, collected from previous attempts to measure anxiety, stress, neurotic debility, and several other previously established general personality factor dimensions. Most readers will wish only the indication of their nature given by their title and description in the text. Those requiring more detailed descriptions can obtain them from the authors or by reference to other researches, where the rationale for selection of personality factor markers is also explained (Cattell, 1955b; Cattell & Scheier, in press-c; Scheier & Cattell, 1958). The main aim in the choice of variables was to make possible the distinguishing of the threat-response factors from other known personality factors.

The environmental conditions, i.e., the stimuli, were varied in a two-step, present-absent, on-off manner. The "on" phase for each is as follows.

Anticipation of Treadmill Run.[5] Just before the testing session began, *S* was given a notice of an appointment, falling within two or three days, to take the "all-out treadmill run," as part of his requirement for the testing program. This test, which *S*s actually took, involved running on a treadmill to the limit of endurance (some had to be helped off the instrument) before an audience of physical education specialists, including at least one woman. *S* was advised of these details in the notice.

Imminence of Academic Examinations. Each *S* was tested once just before semester exams (or in the midst of them)—the "on" stimulus—and once approximately four weeks before examinations, in what we assumed was a relatively carefree period.

Questionnaire "Group Stress Interview" Probe. A list was made of the common themes for generating upset in interview stress (personal defects, economic insecurity, war dangers, threat of illness, etc.). These were given as stimuli in questions which all *S*s answered instead of by interview questions. This loses the freedom to probe where it most hurts for each individual, but it compensates by more standard conditions. Immediately before the test session assigned for "stress-on," *S* was provoked by responding to 40 items reminding him of common sources of worry: e.g., finances, military service, health, academic status, etc. Typical questions were: "Have you ever noticed or suspected any symptoms of insanity in yourself? Specify briefly." "Are you in debt now?" "Are you doing as well in school as your parents (or you yourself) hoped you would? If not, why not?"

It is an important part of the condition-response design (Cattell, 1952) that the various stimuli used be randomized in relation to one another, as in an analysis of variance with several effects. Otherwise, any correlation among the stimuli would have only local meaning—an artifact of the experimenter's design—and would complicate the interpretation of the correlation with the response patterns. Accordingly, a design was set up for combining the high and low (on and off) incidences of the three stimuli with equal frequency. Actually, the essential aim is to have the *changes* from first to second occasion on each stimulus uncorrelated with changes in the other two stimuli, and to have a step down ("on" to "off") between occasion one and two just as frequent as a step up, i.e., to avoid confounding with a particular sequence effect. This requires eight combinations, as shown in Table 1.

A given stimulus was in the "on" condition for approximately half the *S*s at Occasion 1 and "on" for the other half at Occasion 2. All possible on-off combinations of the three conditions occurred at each time; hence, all possible combinations of increments and decrements occurred between Occasion (or "time") 1 and 2. An increment in a given condition ("off" to "on") between Time 1 and 2 was scored in the subsequent correlation as 2, a decrement as 1. Thus for all *S*s in Group (Cell) 1, the stimulus scores entered were T = 2, E = 1, and P = 1; whereas for group (Cell) 2, the corresponding scores entered were 2, 2, and 1. Positive correlation (phi coefficient) between a condition and a response increment variable thus signifies that increase in the amount of this "stress" condition is associated with increase in the response variable, from the first to the second occasion. (No "on-on" and "off-off" sequences were studied in this first attack, though they could be used, giving a middle value in a three score range).

---

[3] On the basis of previous personality profile information on athletes (Cattell et al., 1956), nine varsity athletes were included in the low anxiety group, while the other high and low anxiety subjects were preselected from the upper and lower 10% according to the preliminary criterion of scores on the anxiety factors (second order) in the 16 P.F. Questionnaire (Cattell, 1956). These anxiety selections were not used as criteria, but only to give a good "spread" on anxiety responsiveness.

[4] A good deal of discussion has been devoted by McNemar (1958) and others to the desirability of correcting change scores so that no correlation remains between absolute scores and change scores. Without space to discuss the complex issues involved we shall simply say that (a) this "ideal" seems to us very questionable, though McNemar's formula with the goal of eliminating that part of the correlation which is due to shared error of measurement would be acceptable, and (b) dealing as we are here with changes which are considerable, relative to those occurring in abilities, we have felt it desirable, in a first report, to give results as they emerge with a minimum manipulation of score data, reporting the more refined treatment of error in a later analysis.

[5] Throughout the research, extreme precautions were taken to prevent *S*s from discussing their research experiences with one another. It is believed these precautions were successful, specifically in preventing "anticipation of treadmill stress" at occasion 1, for those *S*'s to whom it was not applied until occasion 2.

TABLE 1

INCIDENCES OF THE THREE "STRESS" STIMULI AT TIME 1 AND TIME 2

Subject[a] Group	Time 1	Time 2 (4 Weeks later)
1	T−, E+, P+	T+, E−, P−
2	T−, E−, P+	T+, E+, P−
3	T−, E−, P−	T+, E+, P+
4	T+, E−, P−	T−, E+, P+
5	T+, E+, P−	T−, E−, P+
6	T+, E+, P+	T−, E−, P−
7	T+, E−, P+	T−, E+, P−
8	T−, E+, P−	T+, E−, P+

Note.—"T," "E," and "P" signify Treadmill, Examination, and Questionnaire Probe Stress. "+" indicates the "on" state, "−" the "off" state.

[a] Due to differential attrition of subjects in the eight groups, *N*'s were only approximately equal for these categories, varying between 10 and 12 for all groups except groups three and six, which had *N*'s of 8 and 14 respectively.

The condition-response design, like analysis of variance, results in a rapid multiplication of the number of categories required (treatment subgroups) as the number of stimuli increases. Consequently, one will usually be constrained to deal with far fewer stimuli than responses, but this suits the objective of the design, which is to examine directly the relation between changes of intensity in single stimuli and changes in intensity of demonstrated response *patterns*. Of course, more response variables are needed to define the latter.

In the classical univariate design, one examines the significance of the difference of the dependent variable values corresponding to the two levels of the stimulus (or the two different effects). The equivalent of this in the present design is to examine the significance of the correlation of the stimulus variable with the response factor.[6] Parenthetically, the *loading* of the stimulus on the response factor, as immediately obtained by the present proposed factoring, is not identical with the correlation when factors are oblique, and requires further calculation; but the departure is so slight with the small obliquities obtained here that we have not considered the complex calculations justified and shall speak of the loading as the correlation.

The condition-response design gives more information than does analysis of variance (*a*) in tying down the "dependent variable" in terms of response dimensions (in groups of tests) and also (*b*) because it reveals not only whether the relation of stimulus change and response change is significant, but also how large it is. On the other hand, it says nothing about interaction effects. However, it should be noted that even if each of the several stimuli were actually perfectly related to the response (in the absence of other variables), the correlation of each therewith could not be unity. If they vary independently (see Table 1 and discussion), the summed squares of their correlations with the common-response could not exceed unity. Thus, in the present instance the three individual loadings could not exceed $\sqrt{.3}$, i.e., .58. Since our experimentally controlled, deliberate manipulation of the stimuli is not the sole variance in these stimuli, the maximum is actually lower than this. For example, if our imposed range of examination stress, between stressless periods of the semester and semester exams, is, on the average, half of the individual's actually experienced range of examination stress, a full response of the anxiety factor in our experiment to the controlled stress would still yield a correlation of only about .41. These points should be kept in mind in our subsequent examination of the stimulus loadings.

RESULTS

Replication of the Four Threat Response Patterns. According to the above design, difference scores on all variables between the first and second occasion, including the prearranged differences on the stimuli, were computed over all 86 individuals, and product-moment intercorrelated. A centroid factor extraction from the 72 × 72 matrix was terminated at 14 factors, according to criteria of completeness of extraction described elsewhere (Cattell, 1952) and rotated to a simple structure which, after 25 rotations, reached a plateau of 59% of variables in the hyperplane.

A separate report gives full data and general descriptive comments on all fourteen factors (Cattell & Scheier, 1958b). The four factors chosen for concentrated attention here as "threat response" factors were so chosen because (*a*) their loaded response variables most strongly suggest the character of "response to threat" as commonly conceived and (*b*) with the minor possible exception noted below, they were the only well identified factors significantly associated with at least one of the three "stress" stimuli. Of the other ten factors, five had a poor or only fair identification with previously known personality dimensions (Cattell, 1957a; Cattell & Scheier, 1958b), and three, though well identified, had no significant stimulus association (U.I. 34, Autistic Nonconformity; U.I. 27, Apathy; and U.I. 35, Self-Reliance). In fact, only two of the ten factors falling outside the original conception of threat-response factors had reasonably good identification plus significant stimulus association. These were factor U.I. 19, Critical Practicality, increased by anticipation of treadmill run, and factor U.I. 17, Inhibition, decreased by examination imminence and by anticipation of treadmill run. If their identification holds up and improves, these two factors may in the future be added to the four threat-response factors, in spite of the fact that the responses they involve are not so obviously "face valid" as are the four factors discussed here.

---

[6] It is unnecessary, of course, to take the correlations of the stimulus change with the changes on the variables in the response factor and to estimate therefrom the correlation with the factor behind them. Factor analysis gives directly the correlation with the true factor.

The four factors which replicate previous patterns, considered as responses to stress in the generalized sense, are: (*a*) Anxiety, recorded as P.U.I. 9(−) as a state and U.I. 24 as a trait (Cattell, 1957a; Cattell & Scheier, 1958a); (*b*) Neurotic Debility, recorded as U.I. 23 as a trait (Cattell, 1957a; Cattell & Scheier, in press-b) and P.U.I. 8(−) as a state (Cattell, 1955a); (*c*) Frustrative Depression or Pathemia vs. Elation, recorded as P.U.I. 2 (Cattell, 1957a) and having resemblance to the Trait Corticalertia vs. Pathemia, U.I. 22, and (*d*) Stress response, P.U.I. 4. The positions of these factors in rank order of diminishing variance of the fourteen factors in this study were 12, 3, 2, and 11 respectively. For ease of reference to the total matrix (Cattell & Scheier, 1958b), these reference numbers will be employed here.

In a moment we shall examine the significance of the relation of change on these factors to stimulus changes, but first let us establish more definitely the nature and identification of the four response patterns themselves. As usual, each pattern is set out in terms of the highest-loading variables with their loadings—actually the *correlations with the reference vectors*, which are proportional to the factor loadings (Cattell, 1952). For economy of presentation, the salients are cut off at a lower loading, one of .20, which is still well above the $P = .01$ significance level for a loading according to Saunders's formula (Cattell, 1952), which indicates a $P = .01$ value of about 0.16. Correlation coefficients, to which these loadings are roughly equivalent, attain significance at the $P = .05$ level, at a value of .21 for an N of 86 cases.

Factor 12 clearly identifies with the Anxiety factor, U.I. 24 (Cattell & Scheier, 1958a), having in common with the established pattern such variables as the symptom check list, the Q4 and C(−) questionnaire factors, and objective test markers[7] (Table 2, Loadings No. 3 and No. 7). Its identification with P.U.I. 9(−), Anxiety state or "high general autonomic activity level" (Cattell, 1955a) is made through the pattern of higher systolic blood pressure (No. 6), lower skin resistance (No. 11), and (at lower loading than the cut-off) more acid pH saliva. Loading Number 12 fits this latter identification only with special assumptions.

Some differences are observable between this pattern of "anxiety as a state" and the earlier studies fixing the pattern of "anxiety as a trait" (*R* technique, U.I. 24). Notably, such U.I. 24 objective test markers as "tendency to agree," "tendency to confess faults," and "severity of criticism" load less here (below .20) than in the trait pattern, though still in the correct direction. The same is true of questionnaire factor O, or Guilt Proneness, and it is tentatively concluded that these are mainly characterological, not *state*, concomitants of anxiety. Conversely, such variables as systolic blood pressure have not appeared reliably in the *trait* pattern.

The second response pattern to be identified —Neurotic Debility—is shown in Table 3. Variables (No. 2) and (No. 4) are those found most frequently to characterize the original Neurotic Debility Factor, U.I. 23(−) (Cattell & Scheier, in press-b); questionnaire factor C(−), lower Ego Strength, also belongs, but the loading for the sway suggestibility marker is only .10, so the pattern of the state must again be considered as somewhat modified from the trait, though still unmistakably characteristic. Less interest in reality-oriented love attachments (No. 1), occurring simultaneously with higher desire to satisfy sexual needs (No. 3), is consistent with a hypothesis of failure to sublimate, but this must be considered speculation pending further research.

---

[7] By "marker," we mean a variable which, through a number of different studies, has significantly loaded a factor in a consistent direction (See Cattell 1955b; Cattell, 1957a). After extensive research has established markers, each factor can be identified in new data in terms of the presence or absence of the group of markers known to load on it from previous work. The markers referred to in the discussion will be most conveniently found in overview references (Cattell, 1955a; Cattell, 1955b; Cattell, 1957a; Cattell & Scheier, 1958a; Cattell & Scheier, in press-a).

TABLE 2

FACTOR LOADINGS ON ANXIETY[a]

1.	+41	More anxiety-tension symptoms (self-checklist)
2.	+40	Q4(+) Greater Ergic Tension
3.	+35	Less confident in assuming skill in untried performance
4.	−32	Less long-circuited (predominance of short term life goals)
5.	−31	Less persistence in unrewarding situation
6.	+30	Higher systolic blood pressure
7.	−30	Lower average handwriting pressure
8.	−29	C(−) Lower Ego Strength
9.	+27	Greater volume of saliva secreted
10.	−26	Less work decrement due to anxiety sound distraction
11.	−26	Lower absolute level of galvanic skin resistance
12.	−22	Smaller pupil diameter during stress
13.	−21	Lower ratio of "acquaintances" to "friends" remembered
14.	+20	More desire to satisfy sexual needs
15.	+20	More slowing of reaction time due to electric shock

Note.—In Tables 2 through 5, we give loadings for all *response* variables which appeared in this study and loaded on (correlated with, associated with) the factor dimensions at .20 or more. Table 6 (left-hand side) gives all *stimulus* associations found, no matter what their size.

[a] Factor 12: U.I. 24(+); P.U.I. 9(−).

TABLE 3

FACTOR LOADINGS ON NEUROTIC DEBILITY[a]

1.	−46	Less interest (objective test) in complex, reality-oriented, love attachments
2.	−40	Lower ratio accuracy (number correct) to accomplishment (number done) on a variety of tasks
3.	+39	More desire (objective test) for sexual pleasure
4.	+30	More rigidity (classical motor and perceptual)
5.	−26	Lower ratio warned/unwarned reaction time
6.	−26	Smaller size of myokinetic movement
7.	+24	More acceleration to competitive situations
8.	−21	C(−) Lower Ego Strength
9.	+21	More errors made in complex reaction time runs
10.	−21	Less reluctance to enter fear situations
11.	+20	More anxiety-tension symptoms (self-checklist)
12.	+20	Prefer weaker (vs. stronger) smells

[a] Factor 3: U.I. 23(−); P.U.I. 8(−).

The original P.U.I. 8(−) state or change-through-time pattern from longitudinal (*P* technique) individual studies (Cattell, 1955a) is recognized also here as a combination of heightened rigidity (No. 4), smaller size of myokinesis (No. 6), and reduced hand tremor—the latter loading below the cutting point. It is concluded that this is the state factor corresponding to the trait of Neurotic Debility, one of the most important factors distinguishing between clinically judged neurotics (U.I. 23(−)) and normals (U.I. 23(+)) (Cattell & Scheier, in press-a). U.I. 23 is empirically independent of Anxiety (Cattell, 1957a; Cattell & Scheier, 1958b), which also significantly discriminates between neurotics and normals (Cattell & Scheier, in press-a).

The next factor has never appeared as an inter-individual difference (*R* technique) or trait pattern, but has appeared in four *P*-technique studies (Cattell, 1957a, p. 651) as a state pattern. The matching rests on "fewer circles used in the cursive miniature situation test" (Table 4, No. 1), reduced myokinetic movement (No. 8), reduced rigidity (No. 9), and loading in the correct direction but below the cutting point, better memory performance and other "concentration" performances, notably low sway, and good multiplying. Cattell (1957a) has hypothesized that this is a *stress* pattern, distinct from anxiety or fatigue factors, like that produced by physical stress or disease, but here (as frequently occurs) produced by mental effort.

The last response pattern, shown in Table 5 cannot be matched with complete statistical confidence to the U.I. 22(−) factor (Low Cortical Alertness; depression and Emotional Immaturity), which it most nearly resembles among known factors. The resemblance extends to a combination of slow reaction time, fewer false reactions in a reaction time test, greater persistence, and (loading below the cutting point) slower tempo, and less fidgeting recorded on fidgetometer. However, there are also several variables present here, notably No. 1, which have never before been found to load U.I. 22. Therefore, although no possible alternative U.I. 22 match exists in the full series of 14 factors in this study, we have contingently named this large variance factor "Low General Excitation Level" as well as "Low Cortical Alertness" or "Pathemia" on the chance that these are real alternatives and not just different names for what is essentially the same functional unity. Conceivably, the alignment with U.I. 22 is correct, but the pattern of the *state* and of the *trait* (relatively permanent high vs. low Cortical Alertness) are more divergent than state and trait commonly are.

TABLE 4

FACTOR LOADINGS ON EFFORT STRESS,[a] OR GENERAL ADAPTATION SYNDROME (STAGE ONE)

1.	−36	Fewer circles used in Cursive Miniature Situation (CMS)
2.	−35	0(−), More confidence, Self-assurance
3.	+30	Overestimate length of fear periods
4.	−26	Lower systolic blood pressure
5.	+24	Greater number of threatening objects perceived in unstructured drawings
6.	−24	Less desire (objective test) for sexual pleasure
7.	+23	More persistence of behavior in unrewarding situation
8.	−21	Smaller size of myokinetic movement
9.	−21	Lower rigidity (Classical motor and perceptual)
10.	−20	Lower index of carefulness

[a] Factor 11: P.U.I. 4(+).

TABLE 5

FACTOR LOADINGS ON PATHEMIA[a] (FRUSTRATIVE DEPRESSION OR LOW CORTICAL ALERTNESS)

1.	+46	More honesty in admitting common frailties
2.	+43	High fear (accident-disease)
3.	−43	Smaller amount of GSR conditioning
4.	−37	Relatively less fluent about self than about others
5.	−34	Fewer errors in complex reaction time
6.	+34	More slow (irregularly warned) reaction time
7.	−34	Smaller magnitude of hand tremor
8.	−33	Less involuntary muscle tension in right arm
9.	+31	More persistence in unrewarding situation
10.	+30	More recovery of GSR level after relaxation from shock
11.	+28	More desire for self control
12.	+27	More threatening objects perceived in unstructured situation
13.	+27	Greater volume of saliva secreted
14.	−26	Fewer circles used in Cursive Miniature Situation test
15.	−26	Less desire (objective test) for sexual pleasure

[a] Factor 2: P.U.I. 2; U.I. 22(−).

The Relation of Response Patterns to Stimulus Conditions. Correlations between the factors (actually, reference vectors) representing the response patterns are given in the right hand section of Table 6. The left hand section of Table 6 gives relations between each of the three stimuli and the four threat-response factors.

These results may be evaluated in terms of (*a*) the significance of a correlation on 86 cases, which must be .21 or above for significance at the $P = .05$ level, or (*b*) the significance of a loading, which is taken here as .16.

TABLE 6

CORRELATIONS OF RESPONSE FACTORS WITH STIMULI AND WITH ONE ANOTHER

	Exam Imminence	Probing Questionnaire	Anticipation of Treadmill Run	U.I. 24(+)	U.I. (23)(−)	P.U.I. 4(+)
U.I. 24(+) Anxiety	−25	00	+13			
U.I. 23(−) Neurotic Debility	+04	+06	−35	+26		
P.U.I. 4(+) Stress or General Adaptation Syndrome	+16	+28	−16	−12	+09	
U.I. 22(−) Lower Cortical Alertness or Excitation Level Lower	+20	+02	−07	+11	−04	+15

We shall summarily depend on Criterion *a* to be on the safe side. On this criterion, it becomes evident, first, that *no single response factor behaves in the same way to all three threat situations*; that is, no response dimension, as a "dependent variable," is significantly affected in the same direction by all three "stress" stimuli. There is not even a consistency in direction of sign for all three stress stimuli on a given factor. Secondly, two of the three highest relations are in the opposite direction to that which naive views might hypothesize. Thus, the anticipation of a treadmill run significantly *reduces* Neurotic Debility and the incidence of examinations significantly *reduces* Anxiety. However, presence of probing questionnaire threat is significantly associated with increased Effort Stress Reaction, P.U.I. 4(+), as would commonly be expected.

DISCUSSION

Although it is a possibility that future research will discover "stressful" stimuli which do act on the Anxiety factor in the naively expected direction, the fact we have to assimilate here is that the expectation can be quite wrong. Experimenters have assumed anxiety to arise from stress stimuli without regard to the clinically common observation that a personality beset by vague fears and anxieties often loses these anxieties for a while when a real danger threatens. For example, in discussing the case of Helen, May (1950, p. 332) states:

> So long as her apprehension could be attached to the possible sufferings in childbirth—a focus of fear which could easily be construed by Helen as 'rational'—she was not faced with the much more difficult problem of confronting the underlying guilt feelings, even the admission of which would have threatened her central psychological protective strategies and thrown her into profound conflict.

Incidence of examinations may be a peculiarly efficacious "focus of fear" in furnishing a pretext for the excessive busyness which is a commonly accepted means of allaying anxiety (May, 1950, p. 172). This phenomenon is not necessarily merely a switch from free floating anxiety to a delimited phobia; a more fundamental process may be involved, which is beyond the space of the present article to discuss. On this point we shall content ourselves with a simply stated hypothesis, arising from the present research and susceptible to checking by a similarly designed study: namely, *level on the anxiety state factor (P.U.I. 9[−]) declines when an external challenging situation, demanding and permitting action, actually materializes after a long period of anticipation.* In fact, the relation of anxiety to alleged anxiety-stimulating situations is temporally much more complex than has been assumed in validating a priori anxiety tests. Thus, here the impact of an actual examination seems to reduce anxiety, while elsewhere (Cattell & Scheier, in press-a) we have shown that news of failure in an examination (where nothing can be done immediately) significantly raises the level on this same defined anxiety factor.

Anticipation of the treadmill run, though it slightly (barely significantly) raises Anxiety, acts powerfully to *reduce* the state we have called Neurotic Debility (see Table 6). The hypothesis that can best be formulated from the present findings is that *challenge to effort, with minimal involvement of major sentiment systems and their attendant anxieties, may slightly raise anxiety, but stimulates the capacity to mobilize (reduces level on the factor of Neurotic Debility (U.I. 23[−]).* The implication favoring therapy involving shock or mild "stress" may be better discussed after the significant finding here is verified with other stimuli relevant to the hypothesis.

Although the results for the effort stress factor (P.U.I. 4) are not all sufficiently significant, it comes nearest to being confirmed as a general, constant, immediate response-to-threat dimension. In fact, if we recall that the positively associated examination incidence and probing questionnaire "threats" are stimuli to which response is already substantially in action at the time of the testing, whereas the treadmill is only something that is *going* to happen, we may tentatively hypothesize that the *Effort Stress reaction pattern, P.U.I. 4[+], appears typically in immediate response to a threat situation demanding effort.* This hypothesis predicts that treadmill stress would have loaded positively on P.U.I. 4 if our response measurements had been made *during* the treadmill run or closer in time to it.

To justify our emerging hypothesis about the fourth of our state factors requires collation from further afield than present space permits. For various reasons, the term "Pathemia" has been applied to this state (U.I. 22(−), P.U.I. 2), describing a state of "depression-with-emotionality" and reduced cognitive clarity characteristically resulting from certain kinds of frustration. In the present experiment, the actual examination situation produced an increase in this pathemic state, but neither of the other stimuli did. We hypothesize that the pathemic response is a depressive response to a frustrating situation too complex for pugnacity, *the depression characteristically showing regressive emotionality along with the depression.*

In general the results support the methodological argument set out elsewhere (Cattell & Scheier, in press-a) regarding the study of states and moods. The first, necessary step is to recognize the chief independent *response* patterns. For what is to be called a stressful, anxiety-provoking, depressing, etc., stimulus, has to be decided by its demonstrable connection with such reactions, not by the experimenter's fiat. Second, the psychological meaning of any stimulus can be defined by its profile of correlations with these independent responses, as in Table 6. Here, this is a statement for the general population, but individuals characteristically show unique combinations of the common factor responses to any given stimulus.

At the same time, the labels we decide to give to the discovered response patterns must, for semantic reasons, fit the popular adjective applied to the majority of situations by which the response patterns are triggered. Thus, the label "stress response pattern" is properly attached to P.U.I. 4, partly because this response turns out to be most connected with stress situations per se, but also partly because the responses loading the pattern are those theoretically to be expected in an organism in stress. Once it becomes so labelled in general psychology, the right of any stimulus to be named "stress" must henceforth depend on its being a *significant provoker of this particular pattern of response*. Actually, the central feature in this stress turns out to be a demand for powerful *effort*, whereas, popularly, "stress" has been used in the title of articles dealing with anxiety-provoking or simply unpleasant or threatening, flight-evoking stimuli. Similarly, the identity of the Anxiety Factor is defined and given meaning originally through the introspected anxiety in many of the responses found loaded in the pattern. However, this meaning is enriched by the significant relations to various stimuli, e.g., examination threat, etc., even though the direction of the relation requires the hypothesis of a peculiarity of personality dynamics.

If these responses have now been defined as recognizably unique response loading patterns, we can henceforth unambiguously designate stress stimuli as those which provoke the stress response factor (P.U.I. 4) and which, according to our hypothesis, have the quality of evoking a great deal of physical and mental *effort*. Anxiety stimuli will be those which produce changes in the U.I. 24 response pattern or, by our hypothesis, seem to affect or threaten the individual's massive life purposes (sentiment systems). Mobilization stimuli will be those which primarily determine changes along the Neural Reserves or U.I. 23 dimension; by our hypothesis, such stimuli challenge and mobilize rather than provoke anxiety or stress.

SUMMARY

1. A condition-response factor analysis has been applied to three stimulus situations and 69 response variables on 86 young adult males, revealing fourteen response patterns of which four are considered to have most importance as threat-response characteristics.

2. These four "state" response factors found among *increments* of response are recognizably similar (*a*) to three factors repeatedly found by *R* technique or "trait" study of absolute levels; namely anxiety (U.I. 24(+)), Neurotic Debility (U.I. 23(−)), and Pathemia or Low Cortical Alertness (U.I. 22(−)) and (*b*) to factors replicated in P-technique, namely, Anxiety state, P.U.I. 9(−), Pathemia, P.U.I. 2, Neurotic Debility, P.U.I. 8(−), and stress or General Adaptation Syndrome (P.U.I. 4(+)).

3. Significant positive relations are found (*a*) between decline on the Anxiety factor level and being in an examination "stress" period; (*b*) between low level on the Neurotic Debility factor and the anticipation of an all-out treadmill run; (*c*) between higher level on the Stress factor and undergoing the probing questionnaire and examination imminence "stresses"; and (*d*) between examination imminence and lowered cortical alertness (Pathemia or Frustrative Depression). Specific testable hypotheses are proposed to the effect that (*a*) response to real danger may reduce anxiety, (*b*) challenge may reduce Neurotic Debility, and (*c*) nocive, unpleasant, or demanding stimuli, when operating so as to require active response at the time of measurement, may produce stress-factor response.

4. In the interests of avoiding problems created only by terminology, it is suggested that the whole class of unpleasant, threatening, or effort-demanding stimuli sometimes called "stress" be more precisely defined, differentiated, and labeled in terms of their loadings on response dimensions, specifically along the Anxiety dimension (U.I. 24[+]), the Effort Stress dimension (P.U.I. 4[+]), the Neurotic Debility dimension (U.I. 23[−]), or the Pathemia dimension (U.I. 22[−]).

Thus, as part of a general taxonomy of situations, it is proposed that what have loosely been called stress stimuli should be classified according to their discovered profiles of loading on these four (and possibly two or three yet unconfirmed) response factors. Individuality of emotional response to the same stimulus could similarly yield precise quantitative statement in terms of unique factor profiles.

REFERENCES

CATTELL, R. B. *Factor analysis: an introduction and manual for the psychologist and social scientist.* New York: Harpers, 1952.

CATTELL, R. B. The chief invariant psychological and psycho-physical functional unities found by P-technique. *J. clin. Psychol.*, 1955, **11**, 319–343. (a)

CATTELL, R. B. The principal replicated factors discovered in objective personality tests. *J. abnorm. soc. Psychol.*, 1955, **50**, 291–314. (b)

CATTELL, R. B. Second-order personality factors in the questionnaire realm. *J. consult. Psychol.*, 1956, **20**, 411–418.

CATTELL, R. B. *Personality and motivation structure and measurement.* Yonkers, N. Y.: World Book Co., 1957. (a)

CATTELL, R. B. A universal index for psychological factors. *Psychologia*, 1957, **1**, 74–85. (b)

CATTELL, R. B., DAY, M., & MEELAND, T. Occupational profiles on the 16 Personality Factor Questionnaire. *Occup. Psychol.*, 1956, **30**, 1–10.

CATTELL, R. B., & SCHEIER, I. H. The nature of anxiety: A review of thirteen multivariate analyses comprising 814 variables. *Psychol. Rep.*, 1958, **4**, 351–388. (a)

CATTELL, R. B., & SCHEIER, I. H. Factors in personality change: A discussion of the condition-response incremental design and application to 69 personality response measures and three stimulus conditions. Advance publication No. 9, Lab. Pers. Assessmt. Group Behav., Urbana, Ill. 1958. (b)

CATTELL, R. B., & SCHEIER, I. H. *The meaning and measurement of neuroticism and anxiety.* New York: Ronald Press, in press. (a)

CATTELL, R. B., & SCHEIER, I. H. The objective test measurement of neuroticism, U.I. 23(−); a review of eight factor analytic studies. *Indian J. Psychol.*, in press. (b)

CATTELL, R. B., & SCHEIER, I. H. Extension of meaning of objective test personality factors: Especially into anxiety, neuroticism, questionnaire and physical factors. *J. gen. Psychol.*, in press. (c)

HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.

MAY, R. *The meaning of anxiety.* New York: Ronald Press, 1950.

MCNEMAR, Q. On growth of measurement. *Educ. psychol. Measmt.*, 1958, **1**, 47–55.

RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.

SCHEIER, I. H., & CATTELL, R. B. Confirmation of objective test factors and assessment of their relation to questionnaire factors: A factor analysis of 113 rating, questionnaire, and objective test measurements of personality. *J. ment. Sci.*, 1958, **104**, 608–624.

SPENCE, K. W., & BEECROFT, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, **48**, 399–403.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

WOODROW, H. The relation between abilities and improvement with practice. *J. educ. Psychol.*, 1938. **29**, 215–230.

(Received May 2, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

INTERFERING AND FACILITATING EFFECTS OF TEST ANXIETY[1]

BRITTON K. RUEBUSH
Yale University

Interfering effects of test anxiety in children were observed in an experimental learning study reported by Waite, Sarason, Lighthall, & Davidson (1958). In another study, Lighthall, Ruebush, Sarason, & Zweibelson (1959) found that while low test anxious (LA) children improved more than high test anxious (HA) children over a two-year time span on the "test-like" Otis Beta Intelligence Test, the HA improved more over the same period on the "game-like" Davis-Eells Test of Problem-solving Ability. These findings suggest that although test anxiety has interfering effects upon intellectual performance under some conditions, it may have *facilitating* effects under other conditions.

The present study was designed as a test of several predictions derived from a simple theoretical model, formulated to account for the fact that test anxiety may have both interfering and facilitating effects upon intellectual performance. In general, the model holds that the effect of anxiety upon performance is mediated primarily by defensive reactions to the anxiety. This hypothesis is supported by the results of a study in which it was found that the Rorschach performance of HA children was significantly affected by their cautious and dependent behavior in the test situation (Sarason, Davidson, Lighthall, & Waite, 1958b). In particular, the model states that (*a*) defensive reactions such as cautiousness have been strongly overlearned and thus are stable "automatic" responses to the anxiety triggered off by such situations such as tests, (*b*) the effects of anxiety upon performance are mediated by these defensive reactions to the anxiety, (*c*) in problem solving situations where such defensive reactions are an asset, anxiety has a facilitating effect upon performance, and (*d*) in problem solving situations where such defensive reactions are a liability, anxiety has an interfering effect upon performance.

The criterion task selected for the present study provided a problem solving situation in which cautious behavior was expected to facilitate performance. Therefore, three specific predictions were made: (*a*) Highly cautious *S*s do better than low cautious *S*s on the criterion task; (*b*) HA *S*s obtain higher cautiousness scores than LA *S*s; and (*c*) HA *S*s perform more effectively on the criterion task than LA *S*s.

METHOD

Subjects and Research Design

The population from which the children in this study were selected was composed of 280 sixth-grade boys from the 11 elementary schools in the predominantly middle-class community of Hamden, Connecticut. Within a period of approximately two months preceding the selection of the final sample, the entire sampling population received the Yale Test Anxiety and Yale General Anxiety Scales, as well as the Beta Test Form Em of the Otis Mental Ability Tests.[2] Previous reports have described the Yale anxiety scales and given evidence pertaining to their validity (Sarason, Davidson, Lighthall, & Waite, 1958a; Sarason et al., 1958b; Sarnoff, Lighthall, Waite, Davidson, & Sarason, 1958; Waite et al., 1958; Zweibelson, 1956). As the first step in selecting the final sample, *S*s whose test anxiety scores were ≤ 5 were designated as LA, and those whose test anxiety scores were ≥ 15 were termed HA. These cutoff points included approximately the lower and upper quartiles of the total sampling population. Second, each of these LA and HA *S*s whose Otis Beta IQ score fell within the interval 89–122 was then assigned to one of three levels of intelligence: 89–100, 101–110, or 111–122, designated as low, medium, and high levels, respectively. Third, eight *S*s were selected by a table of random numbers from each of the six anxiety-IQ groups. Fourth, each of the groups of eight *S*s was randomly split into two groups of four, one of which was to receive the easy criterion task items first, the other to receive the difficult items first. To sum up, the research design of the present study was a 2 (anxiety levels) × 3 (intelligence levels) × 2 (order of administration of easy and difficult items) mixed factorial arrangement, with four *S*s in each cell.

---

[1] This study is part of a project of which Seymour B. Sarason is the principal investigator and which is being supported by a grant from the United States Public Health Service (M-712). The author wishes to express his deep appreciation to Seymour Sarason, Kenneth Davidson, Frederick Lighthall, and Richard Waite for their interest and helpful suggestions in the design and execution of this investigation.

[2] The author wishes to thank Kenneth Davidson and Jane Maltby for their help in collecting these data.

The within-*S*s effect consisted of a main effect of trials (item difficulty) plus the interaction of trials with the between-*S*s effects (anxiety, intelligence, and order).

Criterion Task

The criterion task used in this study consisted of the 12-item short form of the Witkin Embedded Figures Test developed by Jackson (1956). Diagrams and detailed information about the figures are given by Witkin (1950). Previous studies have demonstrated that the criterion task is well within the intellectual grasp of the population used in the present study (Witkin, Lewis, Hertzman, Machover, Meissner, & Wafner, 1954). The task items were ranked in order of difficulty according to the mean solution time achieved on each item by an independent sample of 26 13-year-old-boys.[3] Ranked from easiest to most difficult and using Witkin's (1950) numbering system, the rank order of difficulty of the items was as follows (mean solution times in parentheses): D-2 (19 sec.); G-1 (23); A-3 (25); C-2 (36); E-3 (58); E-5 (60); C-3 (67); H-1 (72); D-1 (79); C-1 (102); A-2 (104); and E-1 (105).

Procedure

E, who did not know the anxiety or IQ level of the *S*s, worked with them individually in the public schools. After bringing each *S* from the classroom to the testing room, a brief conversation took place during which the establishment of good rapport was attempted. Then the following instructions were given.

> This is like the game in the Sunday papers where you have to find a hidden animal in a picture. Only this time you won't be looking for an animal. This is a test to see how fast you can find a design.

E then gave *S* the practice complex figure for 15 sec., during which he asked *S* whether it reminded him of anything. The complex figure was then removed, and *S* was given the simple practice figure for 10 sec., during which *E* advised *S* to "learn this one carefully" because he would have to locate it in the complicated figure he had just seen. Finally, *E* removed the simple figure and presented the complex figure again, at the same time telling *S* to find the simple figure in the complicated one. After *S* had located the practice embedded figure, he was asked to trace its outline with a blunt instrument so that *E* could see that *S* understood the task. Then the instructions continued as follows.

> This is the way the test will be done. You will be shown the complicated design and then the little simple figure. Finally, you will have to locate the simple figure in the complicated one just as you did. The simple design will always be present in the complicated one, and it will always be right side up. Also, if you forget the simple figure, you may ask to see it again as many times as you wish and it won't count against you since I stop the watch while you're looking at the simple figure. And remember, there may be some more lines crossing the simple figure when it is in the complicated figure. Find the simple figure as fast as you can, since I will be timing you and your grade depends on your time, but be sure that the figure you find has exactly the same size and shape as the one you're supposed to be looking for. As soon as you think you have found the figure, tell me at once so that I can note your time and then show me where it is by tracing its outline. If you are right, I will give you the time when you said you saw it. If you are wrong, you will just go on looking and your guess won't count against you. So remember, tell me as soon as you have found the figure, and then show me where it is. Any questions?

E then presented the items in one of two orders, depending upon the order group to which *S* had been randomly assigned. If *S* had been assigned to order Easy-Difficult (ED), he was given the figures in the order, 1–12. If he had been assigned to order Difficult-Easy (DE), the order of administration was reversed—*S* got the figures in order 12–1.

Each time *S* requested to see the simple figure again, *E* stopped the watch and recorded the time and a figure "S." The watch was not stopped during a guess, although when *S* made a guess the time was recorded, and if the guess was incorrect a "G" was also recorded after the time. If *S* had not located the embedded figure within five minutes, this was noted on the recording sheet, and *S* was then shown the correct location of the embedded figure.

Scoring

Latency. The time in seconds taken by each *S* to solve each figure was recorded and then transformed to logarithms in order to achieve normally distributed item distributions and homogeneity of the variances of the 12 groups. The rationale for this use of logarithmic transformations has been presented by Quenouille (1950). Normality of the item distributions was ascertained by observation and the group variances were found to be homogeneous according to the F_{max} test (Walker & Lev, 1953). The logarithms of each *S*'s latency scores on figures D-2 and G-1, A-3 and C-2, E-3 and E-5, C-3 and H-1, D-1 and C-1, A-2 and E-1, respectively, were summed so that, for purposes of the analysis, each *S* had 6 log-latency scores. In addition, a total log-latency score was obtained for each *S* by summing his 12 log-latency scores.

Number of figures solved. Each *S* was also given a score consisting of the total number of figures solved without regard for the time taken to solve them. Observation of the distribution and homogeneity of variances of these scores revealed no consequential deviation from normality, and thus no need for a transformation of scores.

Cautiousness Index. The instructions preceding administration of the criterion task included statements that *S* could request to see the original (simple) figure as often as he liked and without penalty (he was told the watch would be stopped while he looked). In addition, he was told that guessing was permitted and wouldn't count against him, although he was *not* told that the watch would be stopped during a guess. He was also told that if he made a "wrong" guess, he would be so informed and could then continue his search for the embedded figure. Thus, the task was structured so that a cautious approach in which *S* made sure that he knew the correct response before he made a guess would be rewarded by success (since the watch continued running during guesses). A ratio consisting of the number of requests to see the simple figure divided by the number of "wrong" guesses was computed for each *S*. This ratio is relatively independent of latency because it was possible for *S* to obtain high or low ratios regardless of the amount of time he took to solve the figures. The *S* who obtains a high ratio score is one who asks to see the simple figure many times before he ventures a guess, and although there are elements of both dependency and cautiousness involved in this sort of behavior, the ratio score has been labeled the *Cautiousness Index*, because this term seems to provide the most accurate description of the variable under study. Since these ratio scores were not normally distributed, the logarithmic transformation was again utilized successfully to normalize the distribution so that *t* ratios could be computed.

---

[3] The author is grateful to H. A. Witkin for making these data available.

RESULTS

Analysis of Latency Scores[4]

The results of the analysis of variance of the log-latency scores are given in Table 1. Although the main effects of both Anxiety and linear IQ only approached statistical significance, the Anxiety × linear IQ interaction was found to be statistically significant ($p < .05$). This interaction resulted from the fact that, in comparison with the LA *S*s, the performance of the HA *S*s was superior at the low and middle levels of IQ and inferior at the high IQ level. As a statistical check on this observation, separate *t* tests were computed for the differences between the mean log-latency scores of the HA and LA groups at each level of IQ. The results revealed that the HA performed significantly better than the LA at the low IQ level ($p < .01$) and at the middle IQ level ($p < .03$), whereas the superior performance of the LA at the high IQ level only approached statistical significance ($.10 < p < .20$). The mean log-latency scores for the HA and LA groups at each level of IQ, as well as the results of the *t* tests of the differences between these means, are given in Table 2.

TABLE 1

ANALYSIS OF VARIANCE OF LOG-LATENCY SCORES

	SS	*df*	*MS*	*F*	*P*
Anxiety	3.146	1	3.15	2.23	$p < .15$
IQ	4.781	2	2.39		
Linear	4.624	1	4.624	3.28	$p < .10$
Resid.	0.157	1	0.157		
Order	0.005	1	0.005		
A × IQ	6.843	2	3.42	2.43	$p < .10$
Linear	6,388	1	6.39	4.53	$p < .05$
Resid.	0.455	1	0.455		
A × O	2.024	1	2.024		
O × IQ	2.005	2	1.002		
A × O × IQ	1.767	2	0.884		
Error	50.791	36	1.41		
Within Trials (task difficulty)	51.096	5	10.219	14.45	$p < .001$
T × A	0.765	5	0.153		
T × O	15.847	5	3.169	4.48	$p < .001$
T × IQ	2.542	10	0.254		
T × A × O	2.876	5	0.575		
T × A × IQ	3.279	10	0.328		
T × O × IQ	3.858	10	0.386		
T × A × O × IQ	2.635	10	0.264		
Error	127.242	180	0.707		

TABLE 2

MEAN LOG-LATENCY SCORES FOR ANXIETY—IQ GROUPS

	HA	LA	SE_{Dm}	*t*	*df*	p[*]
L IQ	22.29	25.39	0.79	3.92	14	$<.01$
M IQ	21.65	23.58	0.75	2.57	14	$<.03$
H IQ	22.61	21.34	0.88	1.44	14	$.10 < p < .20$

[*] All significance tests two-tailed.

With respect to the within-*S*s effects listed in Table 1, the main effect of Trials and the Trials × Order interaction were both found to be statistically significant ($p < .001$) and to account for the major portion of the within-*S*s variance. The Trials × Order interaction resulted from the fact that both the LA and the HA *S*s did better on the items which came last in sequence. The Trials effect showed that the criterion task items did differ in level of difficulty. Duncan's Test (Snedecor, 1956) was carried out on the six over-all mean log-latency scores for Trials in order to determine more precisely the extent of the difference in task difficulty level of the criterion task items. Since there was no Anxiety × Trials interaction, the overall MS within was used in computing SE_m for Duncan's Test. The results of the 14 simultaneous comparisons indicated that, with the exception of the comparison of the first pair (D-2 and G-1) with the second pair (A-3 and C-2), each of the six pairs of figures differed significantly in task difficulty

---

[4] The author wishes to express his thanks to Norman Anderson for his helpful advice on the statistical analysis.

TABLE 3
MEAN NUMBERS OF FIGURES SOLVED FOR ANXIETY—IQ GROUPS

	HA	LA	SE_{Dm}	t	df	p*
L IQ	8.88	6.38	0.71	3.52	14	<.01
M IQ	9.50	7.88	0.63	2.57	14	<.05
H IQ	8.62	8.75	0.77	0.17	14	NS
Total	9.00	7.67	0.43	3.09	46	<.01

* All significance tests two-tailed.

from every other pair according to Duncan's Test with the 99% protection level.

Analysis of Number Solved

The *t* testing technique was utilized for testing the difference between the mean number of figures solved by the two anxiety groups, as well as for testing the differences between the means of these groups at each IQ level. The results of the *t* tests are given in Table 3. They show that the HA solved significantly more figures than the LA ($p < .01$), although, as in the analysis of the latency scores, the tests of the differences between the means at each IQ level show that the superior performance of the HA was limited to the low and middle levels of IQ. In addition, and in contrast with the results of the analysis of the log-latency scores, the superiority of the LA *S*s at the high IQ level in the present analysis was not found to approach statistical significance.

Multivariate Analysis of Log-latency and Number Solved

A statistical procedure given by Rao (1952) was used in carrying out a simultaneous multivariate analysis of both log latency and number of figures solved. The only significant effects to result from this analysis were the main effect of Anxiety ($p < .05$) and the Order × IQ interaction ($p < .05$). That is, when both variables are allowed to operate simultaneously in the same analysis, the Anxiety × IQ interaction drops out, and the main effect of Anxiety becomes significant, with the HA showing superior over-all performance. It is apparent that when these two variables are combined in an analysis of this type, the marked superiority of the HA *S*s at the low and middle IQ levels overrides the Anxiety × IQ interaction found in the analysis of variance of the log-latency scores. The Order × IQ interaction resulted from the fact that the high and low IQ groups performed more efficiently when they were given the easy figures first, whereas the middle IQ groups performed more efficiently when given the more difficult figures first.

Analysis of Cautiousness Index (CI)

After collapsing the groups on IQ, the mean log CI score for the 24 HA *S*s was .174, whereas the mean log CI score for the 24 LA *S*s was .056. The *SE* of the difference between these means was .054, and the *t* ratio was 2.19 with 46 degrees of freedom. This ratio is significant at the $p < .02$ level of significance (one-tail). The LA mean log CI score was raised considerably by the fact that one LA *S* obtained a log CI score of 1.061. If the highly deviant score of this LA *S* had been excluded, the mean log CI score for the LA group would have dropped to .011. In any event, it remains clear that the HA group was significantly more cautious than the LA group.

The groups were also collapsed on Anxiety in order to determine whether the three IQ groups also differed in degree of cautiousness. The mean log CI score for the 16 high IQ *S*s was .101; for the 16 middle IQ *S*s, .123; and for the 16 low IQ *S*s, .121. Since the differences between these means were not found even to approach statistical significance, it was concluded that the three IQ groups did not differ in degree of cautiousness.

For purposes of further analysis and comparison, it was decided to define the highly cautious *S* (whether HA or LA) as one with a raw S-over-G ratio[5] ≥ 2 (log S-over-G ratio $\geq .301$) because there was a break in the total distribution of cautiousness scores between the 17 *S*s with raw S-over-G ratios ≥ 2 and the 31 *S*s with S-over-G ratios < 2. According to this criterion of high and low cautiousness, there were 11 HA *S*s and 6 LA *S*s who were defined as being highly cautious. The mean log-latency scores, the mean number of figures solved, and the mean log CI scores of the high and low cautious *S*s—both within and across Anxiety groups—are given in Table 4. It is apparent that the HA highly cautious *S*s solved more figures ($p < .001$) in less time ($p < .001$)

---

[5] This ratio is defined in the section describing the scoring procedure.

TABLE 4

		Log-latency	No. Solved	CI
HA	Cautious	20.928	10.27	0.493
	Not Cautious	23.244	7.92	−0.095
	Mean Diff.	2.316*	2.35*	0.588*
LA	Cautious	21.156	9.67	0.512
	Not Cautious	24.197	7.00	−0.097
	Mean Diff.	3.041*	2.67*	0.609*
Total	Cautious	21.042	10.06	0.502
	Not Cautious	23.721	7.39	−0.096
	Mean Diff.	2.679*	2.67*	0.598*

* Significant at the $p < .001$ level (two-tailed).

TABLE 5
MEAN GA AND TA SCORES OF HIGH AND LOW TA, CAUTIOUS AND NONCAUTIOUS *S*s

	Cautious	Noncautious	Diff.
		HA	
TA	18.2	19.8	1.6
GA	15.7	16.5	0.8
		LA	
TA	3.0	2.6	0.4
GA	10.0	6.3	3.7*

* Significant p at $< .01$ level (22 *df*).

than the HA low cautious *S*s. Similarly, the LA highly cautious *S*s also solved more figures ($p < .001$), in less time ($p < .001$) the than LA low cautious *S*s. Furthermore, inspection of the data in Table 4 reveals that the performance of the 11 HA highly cautious *S*s did not differ appreciably from the performance of the 6 LA highly cautious *S*s, nor did the performance of the 13 HA low cautious *S*s differ appreciably from that of the 18 LA low cautious *S*s. Finally, when the groups were collapsed on Anxiety and IQ, the highly cautious *S*s were found to solve more figures ($p < .001$) in less time ($p < .001$) than the low cautious *S*s.

Since General Anxiety (GA) scores from the Yale General Anxiety Scale for Children were also available for all of these *S*s, the mean GA scores as well as the mean TA scores for the highly cautious and low cautious *S*s within TA groups were computed in an attempt to shed some light upon the fact that, contrary to expectations, six of the low TA *S*s were also highly cautious. *T* tests revealed that the mean GA score for the six highly cautious low TA *S*s was significantly higher than the mean GA score for the 18 low cautious low TA *S*s ($p < .01$), even though their TA scores were not significantly different. On the other hand, the mean GA and TA scores of the highly cautious and low cautious high TA *S*s were not significantly different. These means are listed in Table 5.

DISCUSSION

The prediction that highly cautious *S*s do better than low cautious *S*s on the criterion task is confirmed by the findings. This supports the original analysis of the embedded figures task as one in which cautious behavior is an asset in problem solving.

The prediction that HA *S*s obtain significantly higher cautiousness scores than LA *S*s is also confirmed, consistent with the thesis that (*a*) in the case of the test anxious *S*, the evaluative characteristic of tests has become connected with certain unconscious material involving his early interpersonal relationships and conflicts with important authority figures; (*b*) as a consequence, tests have become equated with danger and thus trigger his anxiety; (*c*) the test anxious child gradually has learned certain methods of defending against the anxiety engendered in him by evaluative situations; (*d*) these defenses include the type of behavior here labeled "cautiousness"; and (*e*) because these defenses have been strongly overlearned, they have become "automatic" responses to the anxiety triggered off by tests.

The prediction that the performance of HA *S*s on the criterion task would be significantly better than that of the LA *S*s is confirmed by the results of the multivariate analysis. However, despite the fact that this analysis accounted for both the control and criterion variables simultaneously, more detailed analysis revealed that the superiority of the HA *S*s was limited to the low and medium level of IQ. For these IQ groups, then, wherever the tasks in a test situation are defined or structured so that cautiousness, as here defined, is a positive asset in solving the tasks; the test anxious child will do well. It is in this sense that the results show that anxiety is not always disorganizing, but that under certain conditions it may be facilitating. It is important to emphasize, however, that the superior performance expected of the HA child in some situations is a "fortuitous" result of defensive reactions to his anxiety over which he has little conscious control, and thus does not at all imply that he will do well in all problem solving areas. That is, in problem solving areas where his stable defensive reactions conflict with the type of behavior required for solution of the problem, they are expected to function as a liability rather than as an asset.

The results of the analysis of the cautiousness scores suggest that cautiousness was indeed the defensive reaction to anxiety which served as a mediating variable between anxiety and performance on the criterion task. It is relevant that the 6 *S*s who were low on TA and highly cautious had significantly higher GA scores than the 18 *S*s who were low on TA and less cautious. This seems to constitute further support for the hypothesized strong relationship between anxiety and cautiousness, since in view of their higher GA scores, it is more likely than not that the six low TA, highly cautious *S*s may also be anxious about tests, although they did not admit it on the TA questionnaire. Sarason et al. (1958b) discussed the problem of differentiating among *S*s who are really not anxious about tests and those who are test anxious, but who cannot or refuse to admit their anxiety. It is also relevant to the present discussion that in a Rorschach study (Sarason et al., 1958b), it was found that the HA asked more questions of the examiner and rejected more cards than did the LA. In other words, their response to the test situation was to become dependent or cautious. However, this kind of behavior was not rewarded in the Rorschach situation. *S* got no help from the examiner and no reward for cautious behavior in terms of success and failure. In the Embedded Figures situation used in the present study, on the other hand, cautious behavior was rewarding because the cautious *S* could ask to see the simple figure again without incurring the disapproval of the examiner (and stop the watch in so doing) and could easily tell when his behavior resulted in success. Thus, with these particular instructions and this type of task, the cautious behavior of HA *S*s facilitates their performance. In summary, the findings of the present study agree with certain past findings which support the generalization that the effect of anxiety on performance, whether facilitating or interfering, is mediated primarily by defensive reactions to the anxiety. It follows that two research problems of major importance are (*a*) the detection and detailed theoretical and experimental analysis of common defensive reactions to anxiety and (*b*) the detection and detailed theoretical and experimental analysis of the type of response requisite to success in a wide variety of task-instruction situations. Such research would improve our ability to predict the problem-solving behavior of children with different personality structures in various types of problem-solving situations.

The significant interaction between Anxiety and IQ yielded by the analysis of variance of the log-latency scores was not predicted in advance. The interaction resulted from the fact that the performance of the HA was superior to that of the LA at the low and middle levels of IQ and inferior at the high IQ level. One tentative explanation is that the LA *S*s who were highly intelligent were able to perceive that this situation called for a specific type of approach and performed accordingly, whereas the less intelligent LA *S*s were not this analytic. To the extent that this is valid, the low and medium IQ HA *S*s, in contrast to their LA counterparts in intelligence, have an initial advantage in a situation in which cautious behavior is an asset because they characteristically tend to react with caution in *all* evaluative situations in order to defend against their anxiety.

It is well known that the behavior of two *S*s may appear to be the same in a specific situation although the behavior is a neurotic symptom in one case and not in the other. Just as close observation of such cases over time enables us to differentiate the neurotic behavior from that which is more flexible, it is possible that the type of cautiousness utilized by the LA highly intelligent group may be qualitatively different from that which characterizes the behavior of the HA group. That is, whereas cautiousness *as an anxiety reaction* may dominate all aspects of the response to the anxiety-arousing situation, regardless of appropriateness, the LA child may be able to limit this type of behavior to those situations which he can recognize as requiring this type of response; and, in contrast with HA *S*s at all levels of IQ, he may also be able to combine speed of response with his cautious approach. In any event, this particular aspect of the findings suggests that the differential effects of anxiety upon performance—as mediated by defensive reactions—may vary systematically depending upon both intelligence level and type of task and instructions.

The statistically significant Trials × Order interaction found in the analysis of variance of the log-latency scores is a reflection of the fact that both the HA and LA performed much better on whatever came last in sequence, regardless of order of task difficulty. This result is interpreted to mean that there was a large learning effect operating within trials, with both the HA and LA learning with about equal facility.

The fact that no Anxiety × Difficulty interaction was found in the present study is relevant in at least two respects. First, it is interpreted as supporting the hypothesis that with respect to the effect of anxiety upon performance, qualitative differences in type of task, task instructions, and stable defensive reactions to the anxiety are more important than quantitative differences in the level of difficulty of the tasks. Second, the absence of such an interaction—despite the fact that the analysis also showed clearly that the variation in task difficulty was extremely large—may be of theoretical significance in view of the fact that those who conceive of anxiety as a simple "motivational drive" rely upon the presence of an interaction between anxiety and task difficulty for support of their theory (Castaneda, Palermo, & McCandless, 1956).

Finally, in view of the success of the Yale Test Anxiety Scale for Children in differentiating between efficient and inefficient performers at several levels of intelligence and between high and low cautious *S*s, the results contribute to the validity of the scale for predicting the problem solving behavior of groups. However, it remains clear that the size of the group differences is not so great that one may confidently predict the behavior of individual *S*s.

SUMMARY

Two anxiety scales and an intelligence test were administered to 280 sixth-grade boys. Forty-eight *S*s, divided into 12 groups in a 2 (high and low anxious) × 3 (high, medium, and low IQ) × 2 (order of administration of easy and difficult items) design, were individually administered an embedded figures task. The task items ranged from extremely easy to extremely difficult. Three scores were derived from the criterion task for each *S*: a latency score, number of figures solved, and a cautiousness index. Three predictions were made: (*a*) Highly cautious *S*s do better than low cautious *S*s on the criterion task; (*b*) HA *S*s obtain higher cautiousness scores than LA *S*s; (*c*) The performance of HA *S*s on the criterion task is superior to that of the LA *S*s. The first two predictions were confirmed. The third prediction was confirmed for *S*s at the low and medium IQ levels. The analysis of variance of the latency scores yielded a significant Anxiety × linear IQ interaction which resulted from the fact that the performance of the HA was superior to that of the LA at the low and middle levels of IQ and inferior at the high IQ level. The results of the analysis of the cautiousness scores were interpreted as supporting the thesis that the effect of anxiety on performance, whether facilitating or interfering, is mediated primarily by defensive reactions to the anxiety. It was also concluded that the differential effects of anxiety upon performance may vary systematically depending upon both intelligence level and type of task and instructions.

REFERENCES

CASTANEDA, A., PALMERMO, D. S., & McCANDLESS, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.*, 1956, **27**, 327–332.

JACKSON, D. N. A short form of Witkin's embedded figures test. *J. abnorm. soc. Psychol.*, 1956, **53**, 254–255.

LIGHTHALL, F. F., RUEBUSH, B., SARASON, S. B., & ZWEIBELSON, I. Change in mental ability as a function of test anxiety and type of mental test. *J. consult. Psychol.*, 1959, **23**, 34–38.

QUENOUILLE, M. H. *Introductory statistics.* London: Butterworth, 1950.

RAO, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. F., & WAITE, R. R. A Test Anxiety Scale for children. *Child Develpm.*, 1958, **29**, 105–113. (a)

SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. F., & WAITE, R. R. Rorschach behavior and performance of high and low anxious children. *Child Develpm.*, 1958, **29**, 277–285. (b)

SARNOFF, I., LIGHTHALL, F. F., WAITE, R. R., DAVIDSON, K. S., & SARASON, S. B. A cross-cultural study of anxiety among American and English school children. *J. educ. Psychol.*, 1958, **49**, 129–136.

SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Iowa State Coll. Press, 1956.

WAITE, R. R., SARASON, S. B., LIGHTHALL, F. F., & DAVIDSON, K. S. A study of anxiety and learning in children. *J. abnorm. soc. Psychol.*, 1958, **57**, 267–270.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt & Co., 1953.

WITKIN, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, **19**, 1–15.

WITKIN, A. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, P. B., & WAFNER, S. *Personality through perception.* New York: Harper, 1954.

ZWEIBELSON, I. Test anxiety and intelligence test performance. *J. consult. Psychol.*, 1956, **20**, 479–481.

(Received July 18, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

THE EFFECT OF "SOCIAL CENSURE" AND "SOCIAL APPROVAL" ON THE PSYCHOMOTOR PERFORMANCE OF SCHIZOPHRENICS[1]

DAVID K. CAVANAUGH[2], WALTER COHEN, AND PETER J. LANG[3]
University of Buffalo

IN RECENT years many investigators have been concerned with the effects of special motivational conditions on the psychological deficit displayed by schizophrenics. Some workers have introduced into the experimental situation stimuli which may be described as rewarding; others have utilized socially or physically aversive stimuli. All of these conditions have resulted in changes in performance, and special claims have been made for the effectiveness of one or the other of the motivational conditions. Thus, Pascal and Swenson (1952), Cohen (1956), Cavanaugh (1958), and others (Buss, Weiner, & Buss, 1954; Lang, 1959) report a decrease in psychological deficit under conditions of punishment. Conversely, Stotsky (1957), Peters and Jenkins (1954), and others report that rewards are effective in reducing the schizophrenic's impairment. Rodnick and Garmezy (1957) have conducted an extensive research program designed to examine the effect of social punishment and reward on the performance of schizophrenic *S*s. These authors conclude from their research that even "a minimal punishment" has a deleterious effect on the performance of schizophrenics on a wide range of tasks. An alternative interpretation of their data is that it indicates simply that punishment is an effective modifier of schizophrenic behavior. The direction of the modification is not always in a negative direction. When a punishment is of a general noninformational nature, deterioration is most readily observed. When punishment, however, is associated with specific kinds of responses, a reduction in the frequency of those responses is observed. In Rodnick and Garmezy's research, it is possible to reinterpret some of their results as not necessarily indicating decrement. The arbitrary evaluation of the reduction in responses as improvement or decrement is a function of the specific task and measures utilized by the experimenter. The studies that suggest that punishment may result in improvement in both psychomotor and concept formation tasks use an escape paradigm, whereas most of the studies cited by Rodnick and Garmezy (1957)—where punishment resulted in decrement—use an avoidance design.

In recent research by Lang (1959) and Buss (1954), information was conveyed to schizophrenics about the inadequacy of their responses. In this avoidance situation, the schizophrenics' performance improved. However, in this research the stimulus used to convey information was neutral rather than noxious. The question that arose from consideration of the several studies was whether a noxious stimulus which was presented to *S* (with useful information) would lead to improvement, as in Lang's experiment (1959), or to decrement, as with Rodnick and Garmezy (1957).

EXPERIMENT I

In order to examine these propositions in the present study, information about an inadequate response was introduced into a reaction time situation in two ways. One group of schizophrenic *S*s was presented with a neutral tone whenever the *S*'s response was "too slow." A second group was verbally censured when they responded slowly. On the basis of previous experimental findings and a critical evaluation of the Rodnick and Garmezy material, the following hypotheses were developed: (*a*) *S*s who receive information about their responses under neutral conditions show greater improvement in reaction time than *S*s who do not receive such information; (*b*) *S*s who receive information about their responses under conditions of censure show greater improvement in reaction time than *S*s who do not receive information; (*c*) *S*s who receive information about their performance show greater improvement than those who do not.

---

[1] This study is part of a research program supported by United States Public Health Service Grant M-2173, "Motivation and improved performance in schizophrenia."

[2] Chief Psychologist, Buffalo State Hospital.

[3] Now at the University of Pittsburgh.

Procedure

*S*s for this experiment were 36 male, chronic schizophrenic patients, selected from the in-patient population of the Veterans Administration Hospital, Canandaigua, New York.[4] They ranged in age from 26 to 40 years (mean = 33.64), and all had been hospitalized more than one year. All *S*s had attended high school for at least one year (mean years education = 12.11). No *S* was retained in the sample who had a history of primary mental deficiency or organic cerebral disease.

A disjunctive reaction time apparatus described elsewhere (Lang, 1959), provided the experimental task. *S*s were instructed to respond to the illumination of one of two neon lights by executing a 10.5 in. "jump" with the hand to one of two standard telegraph keys. The "jump" interval was recorded to the nearest .01 sec. by a Standard Electric Timer.

In addition to the lights, sound stimuli were introduced under two conditions of the present experiment. These stimuli were prerecorded on separate tape "loops" and delivered to *S* through earphones from a tape recorder which formed part of the apparatus. The time of presentation and duration of the sound stimuli were controlled by two Hunter Decade Interval Timers.

The initial instructions and procedure were essentially the same as those previously described by Lang (1959). *S*s received a total of 63 trials at the apparatus. These were divided into three series of 21 trials each—preinformation, information, and postinformation. Each series was administered in three blocks of seven trials with a .45 sec. rest between each block.

The preinformation trials were conducted under the usual conditions of administration for a reaction time (RT) task.

On the basis of their performance during this series, *S*s were divided into three matched groups: Control (CO); Tone (TN); and Censure (CN). This initial matching was based on each *S*'s fastest median score for any block of seven reaction times. A matched pairs analysis of variance based on median scores for the 21 preinformation trials revealed that differences between the three groups did not approach significance ($F = .06$, $df = 2$ and 11).[5]

Depending on group assignment, *S*s performed during information trials under one of the following conditions:

1. *TN*. A 300 cps tone was presented when the *S* pressed the correct key, if the *S* responded after a previously established time interval. Depression of the key before the termination of the interval would not activate the tone. The time interval, from onset of the light to the time when pressure on the key would activate the tone, was determined for each block of seven trials by the *S*'s median RT for the previous block of seven trials.

2. *CN*. The verbal message, "That was bad—too slow," delivered authoritatively by a male voice, was presented in the same manner as that described above for the tone. The duration of both the tone and the message was 5.5 sec. Both were presented at low but clearly audible intensity.

3. *CO*. No additional stimuli were introduced into the experimental situation.

Prior to the information trials, *S*s were instructed appropriately, depending on condition assignment. Members of the TN group were told that they would hear a tone through the earphones if they were "too slow" in responding to the lights. The CN *S*s were told that a voice would inform them through the earphones if they were "too slow." To assure the fact that *S*s did not forget the meaning of the voice or tone, they were reinstructed between each block of seven trials that the tone or voice indicated that their previous response had been too slow. These instructions were framed in a nonpunitive fashion. *S*s were told that the voice or tone was designed to help them to respond more quickly by providing information about their performance.

Results

Performance during postinformation and information trials was analyzed in terms of decrement scores. These were obtained by separately subtracting each *S*'s median score for the information and postinformation trials from his median preinformation trials score. The Wilcoxon test for paired replicates (Seigel, 1956) was used to evaluate differences between the three experimental groups.

The results, reported in Table 1, confirmed the previous finding (Lang, 1959): schizophrenic *S*s who receive information about their performance under neutral conditions (TN) show significantly (.05 level) greater improve-

TABLE 1

WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST ANALYSIS OF DIFFERENCES BETWEEN CONTROL AND INFORMATION CONDITIONS

Groups tested	Trials Series	
	Information	Postinformation
CO—TN	11.0[+]	16.0
CO—CN	6.0*	12.0[+]
TN—CN	21.0	29.5

Note.—Entries in this table are the lowest sums by sign (+ or −) of ranked differences between paired replicates. For an N of 12 (as in this case) a sum $<14 = p < .05$.

* $p < .01$.

[+] $p < .05$.

---

[4] The authors are grateful to the professional staff of the VA Hospital, Canandaigua, N. Y. for their assistance in supplying the facilities necessary for the completion of the study.

[5] *S*s who did not achieve a median reaction time for any block of seven preinformation trials of less than .75 sec. were not retained in the sample. Previous research (Lang, 1959) raised doubts that *S*s with slower scores were responding to the experimental instructions.

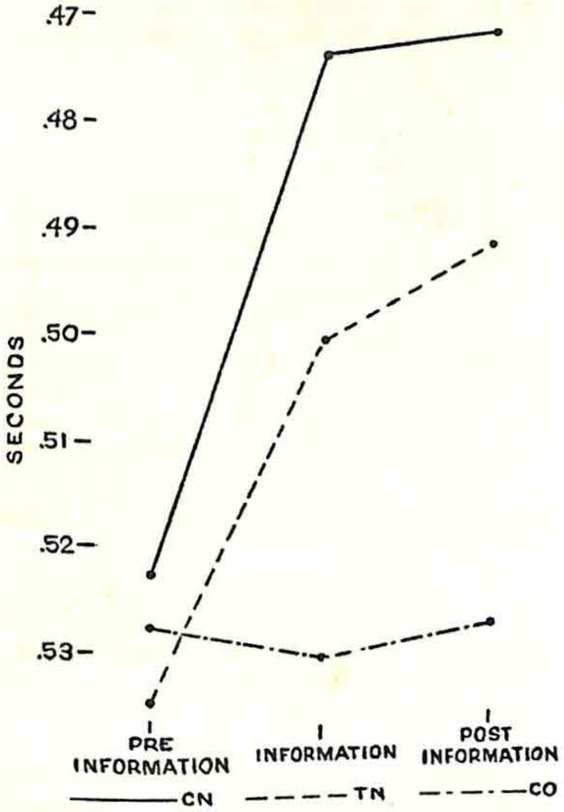


FIG. 1. Mean reaction time for the control and information conditions.

ment than *S*s who do not receive such information (CO). Furthermore, the CN group also improved significantly (.01 level) more than did the control *S*s, and this group maintained the improvement during the postinformation trials. No significant differences were found between the CN and TN groups for either information or postinformation trials. The mean RTs for the experimental groups at each stage of performance are presented graphically in Fig. 1.

EXPERIMENT II

Experiment II was designed to examine the differential effects of social rewards and censure on the disjunctive reaction time (RT) performance of chronic schizophrenic *S*s. Like the Rodnick and Garmezy studies (1957), both the reward and censure conditions were such that *S*s received information about their responses. Three groups of *S*s were employed. One group was informed in a punitive manner that their response had been "too slow" whenever their performance fell below a predetermined score. A second group of *S*s was informed that they were performing well whenever their response latency was faster than a similarly determined standard. A third group received no specific information regarding the adequacy of their performances.

Experiment I indicated that schizophrenic *S*s who receive information about their performance under censure conditions show greater improvement in RT performance than *S*s who do not receive information. The contribution of information in a setting of social approval or reward was considered in Experiment II. It was hypothesized that censure would not be less effective than approval in this setting.

Procedure

*S*s for Experiment II were 27 male, chronic schizophrenic patients selected from the in-patient population of the Veterans Administration Hospital, Canandaigua, New York. They ranged in age from 25 to 39 years (mean = 32.89), and all had been hospitalized more than one year. All *S*s had attended high school at least one year (mean years education = 11.65). No *S* was retained in the sample who had a history of primary mental deficiency or organic cerebral disease.

The apparatus was the same used as that used in Experiment I with the addition of a new tape "loop" for the reward group. A similar procedure was also used. As in Experiment I, on the basis of pre-infomation trials performance, *S*s were divided into three matched groups,[5] (F = .15, not significant for matched pairs analysis of variance of preinformation group means).

During the information trials, *S*s in groups CN and CO (censure and control) performed the reaction time under the same conditions as in Experiment I. In the present experiment the third group, Reward (RE), performed the information trials reaction time under the following conditions: When the *S*'s response was "fast enough," the following message was delivered to the *S* through the earphones: "That was good—very fast." The voice was the same as that used for the punishment condition. The interval during which the striking of the correct response key would lead to the presentation of the message was determined by the *S*'s median reaction time score for the previous seven trials. Whenever *S*'s reaction time was faster than the median, the message was automatically delivered. Whenever *S* was slower than this score no additional information was presented.

The information variable was successfully controlled so that the (CN) and (RE) groups received approximately the same number of messages.[6] Prior to the information trials, *S*s were told about the message they would hear through the earphones, depending on the particular experimental condition. The type of instruction was similar to that used in Experiment I. *S*s were reinstructed between each block of seven trials to assure that they would not forget the meaning of the verbal message. They were told that the voice was intended to help them respond more quickly.

---

[6] The mean number of messages received by *S*s in the experimental groups were 9.8 for the CN group and 10.2 for the RE group.

Results

The difference between a *S*'s median reaction time for the preinformation trials and his score for information trials provided a measure of each *S*'s improvement. Groups were analysed in terms of these scores, using the Wilcoxon test for paired replicates (Seigel, 1956). The results are reported in Table 2. Surprisingly, there was no evidence of difference between the improvement scores of control and reward *S*s for the information series. As in Experiment I, *S*s performing under the CN condition displayed significantly greater improvement than the control group (and also the RE group).

TABLE 2

WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST ANALYSIS OF DIFFERENCES BETWEEN CONTROL AND INFORMATION CONDITIONS

Groups tested	Trials Series	
	Information	Postinformation
RE—CN	3.0+	5.5‡
RE—CO	15.0	14.5
CN—CO	1.5*	3.0+

Note.—Entries in this table are the lowest sums by sign (+ or −) of ranked differences between paired replicates. For an N of 9 (as in this case) a sum $< \sigma = p < .05$.

* $p < .01$.

+ $p < .02$.

‡ $p < .05$.

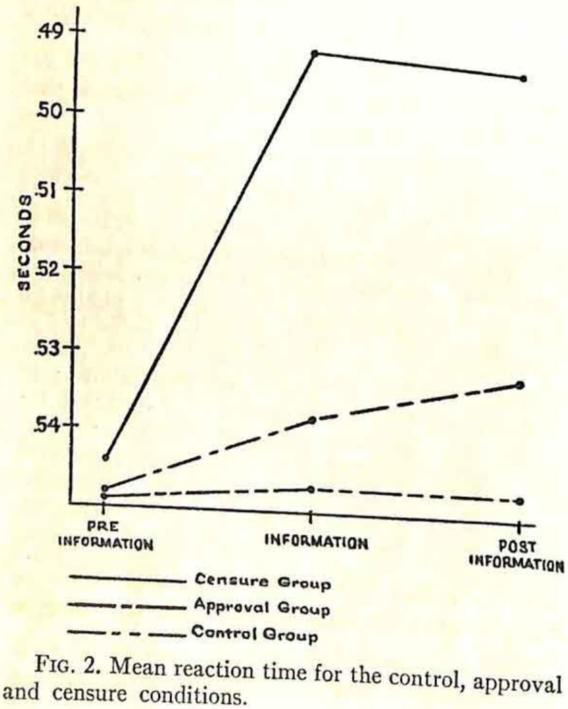


FIG. 2. Mean reaction time for the control, approval and censure conditions.

The performance of *S*s under the postinformation condition were similarly analysed, using as the *S*'s score the difference between median preinformation performance and that achieved during postinformation trials. The CN group again proved to be significantly different from both the Reward and Control groups. The difference in improvement between the CO and RE groups was not significant. The mean RTs for the experimental groups at each stage of performance are presented graphically in Fig. 2.

DISCUSSION

The results of these experiments suggest that social censure can be an effective positive modifier of schizophrenic behavior. This finding is consistent with Buss et al. (1954) and others (Cavanaugh, 1958; Lang, 1959; Pascal, 1952), but not in agreement with Rodnick and Garmezy's (1957) interpretations. It was suggested in the introduction that even the results obtained by Rodnick and Garmezy could logically be interpreted in some instances as not indicating decrement. A good example of this is found in Garmezy's basic study of stimulus generalization, where he reports that schizophrenics were no different from normal *S*s until social censure was introduced. With censure box) flashing "wrong") the normals presented a steepening response gradient, while no essential change was noted in the response gradient of the schizophrenic *S*s. However, Garmezy's procedure involved a differential reinforcement of "push" and "pull" responses. That is, the only response for which any *S* was censured occurred when the *S* indicated that the most dissimilar tone was the same as the training stimulus. Thus, only "pull" responses were punished. These results indicate that the schizophrenics showed a reduction of pull responses. It is possible to interpret these results as indicating that *S* learned to avoid punishment. However, Rodnick and Garmezy prefer to interpret it in terms of a reduction of discrimination, i.e., a flattening of the generalization curve. Similar methodological questions may be raised with regard to a number of other studies cited by Rodnick and Garmezy (1957).

The present results suggest that it is not necessarily censure, as such, which is crucial to the schizophrenic's change in behavior (i.e., the censure and neutral tone groups were not significantly different in degree of improvement). It would appear that the more important factor was the information about inadequate responses itself, rather than the way it was conveyed.

Bugelski (1956) has stressed the importance of "knowledge of results" in any task in which improvement is to be expected with practice, particularly as this knowledge modifies the *S*'s "self instruction" or "set" to respond. For most tasks, the normal *S* will readily "recognize the correctness or wrongness of his responses as soon as they occur, and no outside agency is required" (Bugelski, 1956). The spontaneous verbalizations of normal *S*s in a previous study (Lang, 1959) suggest that this is valid for the experimental task used in this research. However, the work of a number of writers implies that it is not true for schizophrenic *S*s. Schizophrenics appear to be more distractable (Cameron, 1947; Chapman, 1956), and less able than normals to maintain a set to respond to a class of stimuli (Shakow, 1950). Thus, additional informational cues introduced by the *E* may be of much greater importance to psychotic *S*s. From this point of view, censure, as such, may be less crucial to schizophrenic deficit than aspects of the experimental situation which tend to modify set or distract or focus the *S* on the task.

The results of Experiment II suggest that reward (in an informational sense) may have little tangible effect upon the reaction time performance of chronic schizophrenic subjects. Nevertheless, the inference that reward is never effective with schizophrenics is not a necessary conclusion from the present findings. Stotsky (1957) reported some improvement of RT performance under "reward" (supportive urging) conditions with schizophrenics. It is assumed that the present use of reward helped gain the cooperation and interest of the *S*s. In addition, however, it was aimed more precisely at modifying specific responses as they occurred.

It is possible that with a reward condition such as that used in the present design, the function of verbal approval is to confirm an on-going performance without necessarily leading to any change in set. In the censure condition, however, *S* is given a clear indication that he is performing inadequately, and in a sense, instructions to do better are implicit. The effects of these differential qualities of approval and censure, specifically as they modify the *S*'s set to respond, may account for the present findings. This interpretation is consistent with the gradual improvement of the RE group noted in Fig. 1, as opposed to the rapid acceleration of the CN group. Differential implications of reward and punishment for the *S*s' set to respond may also account for the hierarchy of effective reinforcers found by Buss et al. (1954) and for the findings of Rodnick and Garmezy (1957).

In a recent study of reaction time performance with brain damaged patients, Blackburn (1958) reports that *S*s who were informed about their performance and told they could do better improved significantly more than *S*s who were told that they had done well. Although other factors were involved, this finding tends to support the interpretation suggested by the present investigation. Blackburn's results point up the desirability of studies comparing schizophrenic and brain-damaged patients on the present task, as well as tasks used by others in the investigation of deficit and schizophrenia.

In general, the present findings indicate the need for examining rather extensively the specific qualities and implications of the stimuli one labels as rewarding or punishing. Such examination may well benefit from reference to a theoretical framework which includes consideration of the role of reinforcers in modifying and strengthening *S*'s set to respond.

SUMMARY

The present study was designed to examine the effects of information about inadequate responses on the disjunctive reaction time of chronic schizophrenic *S*s. Four experimental conditions were introduced: control, neutral information, censure, and approval. Under the neutral information condition, a "slow" response (based on a previous estimate of *S*'s

performance) resulted in the brief presentation of a 300 cps tone, delivered to *S* through earphones. Under censure conditions, a slow reaction time (relative to previous performance) resulted in the presentation of a verbal message, disapproving of the *S*'s performance. A fast reaction time was followed by verbal approval.

Analysis of the data revealed that both the neutral information and censure conditions led to an improvement in the reaction time significantly greater than that observed among *S*s who did not receive additional information. No significant difference was obtained between the censure and neutral information groups. The approval or "reward" condition had little effect upon *S*s' performances as compared with controls.

The findings were compared with other studies where conflicting evidences regarding the role of reward and punishment were discussed. The findings were interpreted as evidence against the notion that social censure is necessarily productive of deficit. An interpretation of the relative effectiveness of approval, censure, and information was offered based upon the concept of set.

REFERENCES

Blackburn, H. L. Effects of motivating instructions or reaction time in cerebral disease. *J. abnorm. soc. Psychol.*, 1958, **56**, 359–366.

Bugelski, B. R. *The psychology of learning.* New York: Henry Holt, 1956.

Buss, A. H., Weiner, M., & Buss, Edith. Stimulus generalization as a function of verbal reinforcement combinations. *J. exp. Psychol.*, 1954, **48**, 433–436.

Cameron, N. *The psychology of behavior disorders.* New York: Houghton Mifflin, 1947.

Cavanaugh, D. K. Improvement in the performance of schizophrenics on concept formation tasks as a function of motivational change. *J. abnorm. soc. Psychol.*, 1958, **57**, 8–12.

Chapman, L. J. Distractibility in the conceptual performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 286–291.

Cohen, B. D. Motivation and performance in schizophrenia *J. abnorm. soc. Psychol.*, 1956, **52**, 186–190.

Garmezy, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, **20**, 253–276.

Lang, P. J. The effect of aversive stimuli on reaction time in schizophrenia. *J. abnorm. soc. Psychol.*, 1959, **59**, 263–268.

Pascal, G. R., & Swensen, C. Learning in mentally ill patients under conditions of unusual motivation. *J. Pers.*, 1952, **21**, 240–249.

Peters, H. N., & Jenkins, R. L. Improvement of chronic schizophrenic patients with guided problem solving, motivated by hunger. *Psychiat. Quarterly*, 1954, **28**, 84–101.

Rodnick, E. H., & Garmezy, N. An experimental approach to the study of motivation in schizophrenia. In M. Jones (Ed.), *Nebraska Symposium on Motivation.* Lincoln, Nebraska: Univer. of Nebraska Press, 1957. Pp. 109–184.

Shakow, D. Some psychological features of schizophrenia. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950. Pp. 383–390.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Stotsky, B. A. Motivation and task complexity as factors in the psychomotor responses of schizophrenics. *J. Pers.*, 1957, **35**, 327–343.

(Received August 23, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

COGNITIVE AND EMPATHIC FACTORS IN ROLE-TAKING BY SCHIZOPHRENIC AND BRAIN DAMAGED PATIENTS[1]

NORMAN A. MILGRAM[2]
Boston University

THIS study investigates deficiencies in role-taking on the part of schizophrenic patients and nonpsychotic brain damaged patients. Cronbach (1958) and others (Gage, 1955; Bronfenbrenner, 1958) have cited the theoretical and methodological problems encountered in research on role-taking and interpersonal perception. The definitions and operations employed in this study were selected in an effort to avoid such difficulties. Role-taking is defined as an implicit, empathic process whereby a person predicts the behavior in a given situation of another person or persons. According to Taft (1955) role-taking is said to involve two different types of empathy, *individual empathy* and *mass empathy*. The former is an intuitive process by which one experiences the thoughts and feelings of another person. The latter is an inferential process whereby one uses socially shared, conventional frames of reference about groups and group members as a basis for predicting their behavior. The present study is concerned with the analytic judgments required in mass empathy, and investigates the ability of subjects to identify the verbal responses of men, women, adults, and children in multiple-choice word association tests.

Previous studies in this area (Cline, 1955; Dymond, 1950; Sarbin, 1954; Wittick, 1955) have shown a positive relationship between empathic ability and psychological adjustment. A number of theorists (Cameron, 1943; Dunham, 1944; Faris, 1944; Hoskins, 1946; Sullivan, 1944) have suggested that schizophrenia in particular is characterized by a breakdown in role-taking skills. Several studies (Flavell, 1955; Helfand, 1956; McGaughran & Moran, 1956; Sarbin & Hardyck, 1955) have borne out this assertion.

These investigations have not considered, however, the possibility that both *cognitive* and *empathic* abilities are required for an effective role-taking performance. In the typical role-taking test, the subject is called upon to give his own idiosyncratic responses on one form of a preference test and to simulate the answers of another person or persons on the same or a second form of the test. It would appear that the cognitive ability to shift from giving one's preferred or dominant responses to deliberately giving the nondominant responses that characterize another is a prerequisite to simulating the responses of another. A second cognitive ability is the forming of concepts as measured by the ability to select the terms that are normally connoted by a conceptual label, and the rejection of terms that are not. In the present study where *S*s are instructed to choose the verbal responses that men, women, adults, and children characteristically give to certain stimulus words, these *S*s must have, as a cognitive minimum, the ability to attach relevant connotations to these conceptual headings, *men*, *women*, etc.

It has been generally agreed that brain-damaged patients suffer some degree of cognitive deficit (Hunt & Cofer, 1944). With respect to cognitive deficit in schizophrenia, some theorists (Goldstein, 1943) maintain that schizophrenic patients demonstrate cognitive deficits not unlike those observed in brain-damaged patients. Others (Cameron, 1939; Sullivan, 1944) argue that what is lessened or lost in schizophrenia is not cognitive abilities, but the intent or the ability to take the role of the other as a guide to effective communication with others:

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirement for the degree of doctor of philosophy at Boston University, 1958. Sincere thanks are extended to Austin Berkeley, Henry Weinberg, and Harold Goodglass, who served as members of the author's thesis committee. The writer is indebted to C. Bennett and B. Stotsky who contributed to this study by their critical readings and comments. The assistance of the staffs and patients of the Boston and Bedford Veterans Administration Hospitals is gratefully acknowledged. An abbreviated form of this study was presented at the meetings of the Midwestern Psychological Association in May, 1959.

[2] Now at the Nebraska Psychiatric Institute, Omaha, Nebraska.

Hypothesis I. *Schizophrenic patients and brain damaged patients are deficient in role-taking in comparison with normals.*

Hypothesis II. *Brain damaged patients are deficient in cognitive abilities in comparison with schizophrenic patients.*

Hypothesis III. *Schizophrenic patients are deficient in role-taking in comparison with brain damaged patients when differences in cognitive abilities are taken into account.*

METHOD

Subjects

Three male groups consisting of 32 schizophrenic patients, 30 brain damaged patients, and 20 normal *S*s were used.

Criteria for selection of the schizophrenic group included the following: (*a*) all patients had been given a psychiatric diagnosis of schizophrenia, the diagnostic breakdown showing paranoid (24), hebephrenic (2), catatonic (2), and chronic undifferentiated (4); (*b*) no patient had neurological involvement, had undergone somatic therapy (electric shock or insulin coma) within the past six months, or had ever undergone a lobotomy.

Criteria for selection of the brain damaged group included these observations: (*a*) all patients had been given a neurological diagnosis of damage to the cerebral cortex, these diagnoses including vascular disease (9), degenerative disease (12), chronic brain syndrome (9); (*b*) no patient had a previous history of severe mental disturbance or was known to be psychotic at time of testing.

The three groups were selected with attention to age, IQ (as measured by the Vocabulary Subtest of the Wechsler-Bellevue Form I), and education. The data are summarized in Table 1. Differences in age and IQ were not significant. A statistically significant but small difference in years of formal education between normals and the two patient groups was observed.

Role-taking Tests

Multiple-choice word association tests were chosen as measures of role-taking performance because the various responses to each stimulus word could be precoded as to the age or sex group into which they fell. More important, normative data were available as to the empirically observed word association preferences of these groups.

TABLE 1

AGE, IQ, AND YEARS OF FORMAL EDUCATION

Group		Age	IQ	Education
N	*M*	44.6	116.2	12.0
	SD	15.2	10.0	2.7
S	*M*	43.1	114.9	10.8
	SD	13.4	15.7	3.2
BD	*M*	45.6	110.5	11.5
	SD	15.2	17.7	3.2

Terman-Miles Masculinity-Femininity Word Association Test (1936). *S*s were instructed to choose the verbal associations characteristically given by men on one form of the test and by women on the other. Each form contained 60 items. In any given item, there was a stimulus word printed in capital letters and four response words in lower case. One or more of these responses had been empirically demonstrated to be chosen more frequently by men, one or more by women, and the remaining responses were nondiscriminating. Masculine responses were scored +, feminine −, and neutral zero.

Abstract-Concrete Word Association Test.[3] The paradigm for this test was taken from Flavell (1956) and was based on empirical evidence in the literature (Woodrow & Lowell, 1916) and in pilot studies supporting the assertion that adults choose abstractly related associations (synonym, superordinate, and subordinate) in preference to concretely related associations (part-whole, adjective, verb, concrete context) and children do the reverse. Each form of the test contained 48 items. All stimulus words were nouns. Response words were nouns, adjectives, and verbs. In any test item, the stimulus word bore an abstract relationship to one of the response words and a concrete relationship to the other, e.g., DOG—animal (superordinate), bark (verb). The score was the number of abstractly related associations chosen under each of two sets of instructions, i.e., to take adult and child roles.

Cognitive Tests

Mental Set Shifting Test. This test, taken from Appelbaum (1958), involved three presentations of 50 words from the Kent-Rosanoff list (1938). On the first presentation, free associations were obtained; on the second, *S*s were asked to respond with the same word they had given previously; on the third, *S*s were instructed to give responses different from the previous ones. The score was the number of different responses on the third presentation. The rationale for this measure is that *S* is called upon to inhibit his own dominant responses and to adopt and sustain the mental set to give nondominant responses. An analogous cognitive process is required in role-taking.

Concept Forming Test. This test, adapted from Capps (1939), involved a discrimination between relevant and irrelevant associations to a stimulus word. *S*s were instructed to select 3 of 7 response words which went with the stimulus word in some way. The score was the number of items (25 in all) in which all 3 related words were selected. This measure is analogous to a second cognitive requirement of role-taking, i.e., that of selecting the terms that are normally connoted by a conceptual label and rejecting the terms that are not.

---

[3] A copy of the Mental Set Shifting, Concept Forming, and Abstract-Concrete Word Association Tests has been deposited with the American Documentation Institute. Order Document No. 6133, from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies.

Procedure

*S*s were tested individually, usually in one session. The Mental Set Shifting Test was administered first in order to obtain each *S*'s normally dominant responses. Two alternate forms of the Terman-Miles and of the Abstract-Concrete Word Association Tests were administered next with instructions to choose the associations of men and women on the first pair and of adults and children on the second. The Concept Forming Test and Vocabulary Subtest followed.

RESULTS

Hypothesis I. Means and standard deviations of the role-taking scores for the three groups are presented in Table 2. It is evident that the difference between instructions was much greater for the normals than for either of the two patient groups. Since the over-all interaction between Instruction and Groups was significant and in the predicted direction (Table 3), Hypothesis I was confirmed. Comparison of the three groups on the separate scores shows that normals were superior to both patient groups on male scores, but only to the brain damaged group on female scores. Normals were superior to the brain damaged group on both adult and child scores, but were not significantly different from the schizophrenic group on either.

TABLE 2
MEANS AND STANDARD DEVIATIONS OF ROLE-TAKING SCORES

Measure	Group	*M*	*SD*	*D*	p[a]
Male	N	6.60	6.00		
	BD	−.50	6.79	7.10	<.001
	S	−4.28	8.14	3.78	<.05
Female	N	−15.05	7.53		
	S	−11.35	8.87	3.70	.13
	BD[b]	−8.93	9.28	2.42	.28
Adult	N	35.85	6.27		
	S	32.91	6.63	2.94	.11
	BD	28.90	5.56	4.01	<.02
Child	N	20.05	8.00		
	S	23.41	7.72	3.36	.10
	BD[b]	26.23	6.83	2.82	.12

[a] All statistical tests are two-tailed.

[b] Normals were superior to the Brain damaged group on this comparison beyond the .02 level.

TABLE 3
ANALYSIS OF VARIANCE OF ROLE-TAKING SCORES

Source	Male-Female			Adult-Child	
	df	Mean Square	p	Mean Square	p
Groups	2	224.43	—	5.48	—
Subjects	79	81.38		60.84	
Instructions	1	5071.61	<.001	2987.80	<.001
Instructions × Group	2	740.24	<.001	1029.63	<.001
Subjects × Instructions	79	45.19		27.19	
Total	163				

TABLE 4
MEANS AND STANDARD DEVIATIONS OF COGNITIVE TESTS

Measure	Group	*M*	*SD*	*D*	p[a]
Mental Set Shifting	N	45.30	4.41		
	S	42.59	8.80	2.71	.15
	BD	32.46	12.08	10.13	<.001
Concept Forming	N	23.70	1.14		
	S	22.03	3.62	1.67	<.03
	BD	21.73	2.70	0.30	—

[a] The *t* test with Snedecor's Correction for heterogeneity of variance was used (McNemar, 1949). With respect to Concept Forming scores, value of *H* from Kriskal-Wallis analysis of variance scarcely approached significance and the differences between normals and patient groups are of questionable significance.

TABLE 5
ANALYSIS OF COVARIANCE: MENTAL SET SHIFTING[a] ON ROLE-TAKING SCORES

Measure	Source	Before Adjustment			After Adjustment		
		df	Mean Square	p	*df*	Mean Square	p
Male	Between	1	221.58	.05	1	718.41	<.001
	Within	60	56.53		59	43.73	
	Total	61			60		
Female	Between	1	89.96	.25	1	2.00	.89
	Within	60	82.22		59	80.16	
	Total	61			60		
Adult	Between	1	248.52	<.02	1	227.03	<.02
	Within	60	37.46		59	37.89	
	Total	61			60		
Child	Between	1	123.75	.11	1	11.31	.88
	Within	60	45.55		59	43.44	
	Total	61			60		

[a] An Arcsin transformation (Walker & Lev, 1953) was used to make the variances homogeneous as a necessary condition for covariance adjustment.

Hypothesis II. Means and standard deviations of scores on the cognitive tests are presented in Table 4 and show that the hypothesis is confirmed with respect to mental set shifting.

The difference in concept forming between the two patient groups was not significant.

Hypothesis III. Hypothesis III held that the mean role-taking score of the brain damaged group is greater than that of the schizophrenic group when a correction is made for differences in cognitive abilities (i.e., mental set shifting). This assertion was confirmed for male scores only. The data on analysis of covariance of mental set shifting scores and the four role-taking scores are summarized in Table 5. Analyses of covariance for female and child scores revealed no significant differences between the patient groups, although the adjustment was in the predicted direction. With respect to adult scores, the schizophrenic group was superior both before and after covariance adjustment.

DISCUSSION

Hypothesis I. Experimental findings have confirmed the hypothesis that schizophrenic and brain damaged patients are deficient in role-taking performance. These findings are consistent with interpersonal theory (i.e., Cameron, 1939), which states that schizophrenics are deficient in taking the role of the other, and with cognitive theory (i.e., Goldstein, 1943), which states that brain damaged patients do poorly in role-taking tasks because of their deficits in the cognitive abilities required for effective role-taking.

The specific deficiency of the schizophrenic group in taking the male role is also consistent with the psychoanalytic theory of a relationship between paranoid schizophrenia (24 of the 32 schizophrenic patients were paranoid) and feminine identification (Freud, 1924). It is worthwhile noting that the use of the Terman-Miles Word Association Test to discriminate between normals and paranoid schizophrenics has singularly failed until now. Musiker (1952) cites a number of studies, including his own, to make this point. In these earlier studies, the test was administered with instructions "to choose the word that goes best or most naturally with the stimulus word." Male *S*s tended to choose both male and female responses with the latter in the majority by a narrow margin. In the present study where *S*s were specifically instructed to discriminate male and female responses, normals were successful in this deliberate effort, but schizophrenics were not.

Hypothesis II. Results on the Mental Set Shifting Test suggest that the inability to inhibit one's dominant or preferred mode of response is a specific characteristic of thought pathology in organic brain damage. Schizophrenics as a group were not significantly different from normals. The few schizophrenics who did poorly on this test tended to be older, chronically ill patients. Their "brain-damaged-like" performance may have been related either to chronicity of illness, to the mental deterioration that presumably accompanies senescence (Wechsler, 1944), or both. Rank order correlation coefficients between age and mental set shifting were .46 for the brain damaged group and .61 for the schizophrenic group. The higher correlation for the schizophrenic group may reflect a compounding of chronicity of illness and age. The older schizophrenic patients had been hospitalized for many years (up to 40), whereas the brain damaged patients, regardless of age, had all had fairly recent onsets of illness and hospitalization.

It is felt that the Concept Forming Test failed to support the hypothesis because of its low ceiling of difficulty. Even the difference between the normals and the two patient groups was of questionable significance.

Hypothesis III. The magnitude of the correlations which obtained between mental set shifting and male-female role-taking scores confirms the original contention, that the role-taking task has cognitive as well as empathic requirements. The covariance adjustment of the male scores highlights the specifically *empathic* deficiency of the schizophrenics in taking the male role.

Several explanations come to mind to account for the failure of the brain damaged group on the adult-child role-taking polarity. This task implicitly requires a discrimination between abstract and concrete word relationships. Considerable evidence has shown that brain damaged patients are especially deficient in this discrimination (Milgram, 1959); hence, they are penalized by an additional cognitive requirement which does not apply in the male-female role-taking task. Moreover, in the male-female polarity, an *S* is aided in his choices by specific knowledge about the habits, interests, feelings, and attitudes of men and women; on the adult-child polarity, a subject is called upon to identify a subtle quality of thinking that is reflected in language usage. Such a task, requiring a discrimination between the thought characteristics of adults and children, may be especially difficult for the brain damaged group.

At any rate, the performance of the brain damaged group on the adult-child polarity was so poor as to recommend the test as a little understood but highly interesting diagnostic tool. If 7 is taken as the cutoff point for each *S*, then three normals were misdiagnosed because of their failure to shift at least 7 points from adult to child score, whereas only two brain damaged *S*s were misdiagnosed because they shifted more than 7 points.

Implications for future research extend in several directions. The present study should be supplemented by testing the ability of female schizophrenics to take male and female roles. Theoretically, an underlying masculine identification would result in a poor performance on female role. A second study should deal with the adult-child polarity in children. It would be interesting to learn at what age a child follows the adult pattern of preference for abstract relationship in his free associations and in his taking the role of adult. This study would provide a clue to the relationship between role perception and role enactment. Finally, the diagnostic implications of the various cognitive and role-taking tests employed herein merit further study.

SUMMARY

The purpose of this investigation was twofold: to study the nature of role-taking performance in schizophrenic and brain damaged patients, and to contribute thereby to an understanding of thought pathology in these patient groups. Schizophrenic and nonpsychotic brain damaged *S*s were administered four multiple-choice word association tests which required shifting from the male to female role on one pair and from the adult to child role on the other. The latter, for which empirical norms were obtained, required an implicit discrimination between abstract and concrete relationships of stimulus and response words. The most relevant cognitive aspect of the role-taking performance was tapped by a task requiring the ability to inhibit one's preferred response in favor of another.

When role-taking scores were adjusted for differences in this cognitive ability, the brain damaged group was superior to the schizophrenic group on the male-female polarity, chiefly because of the schizophrenics' inability to take the male role. The brain damaged group failed completely on the adult-child polarity to discriminate the abstract from the concrete relationship with the result that the schizophrenics were superior in the shift from adult to child role. Both patient groups were deficient in comparison with normal controls on both role-taking polarities. Results were interpreted as indicating that lowered role-taking performances in schizophrenic and brain damaged groups were related to empathic and cognitive deficiencies respectively.

REFERENCES

APPELBAUM, S. Automatic and volitional processes in the verbal responses of brain-damaged and normal subjects. Unpublished doctoral dissertation, Boston Univer., 1958.

BRONFENBRENNER, U. The study of identification through interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford Univer. Press, 1958, Pp. 110–130.

CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34**, 265–270.

CAMERON, N. The paranoid pseudo-community. *Amer. J. Sociol.*, 1943, **49**, 32–48.

CAPPS, H. Vocabulary changes in mental deterioration. *Arch. Psychol., N. Y.*, 1939, No. 242.

CLINE, V. Ability to judge personality assessed with a stress interview and sound-film technique. *J. abnorm. soc. Psychol.*, 1955, **50**, 183–187.

CRONBACH, L. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford Univer. Press, 1958, Pp. 353–379.

DUNHAM, H. The social personality of the catatonic-schizophrene. *Amer. J. Sociol.*, 1944, **49**, 508–518.

DYMOND, ROSALIND. Personality and empathy, *J. consult. Psychol.*, 1950, **14**, 343–350.

FARIS, R. Reflections of social disorganization in the behavior of a schizophrenic patient. *Amer. J. Sociol.*, 1944, **50**, 134–141.

FLAVELL, J. Thought, communication, and social integration in schizophrenia: An experimental and theoretical study. Unpublished doctoral dissertation, Clark Univer., 1955.

FLAVELL, J. Abstract thinking and social behavior in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52,** 208–211.

FREUD, S. Psychoanalytic notes on an autobiographical account of a case of paranoia. *Collected papers.* Vol. III. London: Hogarth, 1924.

GAGE, N., & CRONBACH, L. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62,** 411–423.

GOLDSTEIN, K. Significance of psychological research in schizophrenia. *J. nerv. ment. Dis.*, 1943, **97,** 492–507.

HELFAND, I. Role-taking in schizophrenia. *J. consult. Psychol.*, 1956, **20,** 37–41.

HOSKINS, R. *Biology of schizophrenia.* New York: Norton, 1946.

HUNT, J., & COFER, C. Psychological deficit. In J. Hunt (Ed.), *Personality and the behavior disorders.* Vol. II. New York: Ronald Press, 1944, Pp. 971–1032.

MCGAUGHRAN, L., & MORAN, L. "Conceptual level" versus "Conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups. *J. abnorm. soc. Psychol.*, 1956, **52,** 43–50.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.

MILGRAM, N. Preference for abstract versus concrete word meanings in schizophrenic and brain-damaged patients. *J. clin. Psychol.*, 1959, **15,** 207–212.

MUSIKER, H. Sex identification and other aspects of the personality of the male paranoid schizophrenic. Unpublished doctoral dissertation, Boston Univer., 1952.

ROSANOFF, A. *Manual of psychiatry.* (7th ed.) New York: Wiley, 1938.

SARBIN, T. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge, Mass.: Addison-Wesley, 1954, Pp. 223–258.

SARBIN, T., & HARDYCK, C. Conformance in role perception as a personality variable. *J. consult. Psychol.*, 1955, **19,** 109–111.

SULLIVAN, H. The language of schizophrenia. In J. Kounin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. of California Press, 1944, Pp. 4–16.

TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52,** 1–23.

TERMAN, L., & MILES, CATHERINE. *Sex and personality.* New York: McGraw-Hill, 1936.

WALKER, HELEN, & LEV, J. *Statistical inference.* New York: Holt, 1953.

WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

WITTICK, J. The generality of the prediction of self reports. *J. consult. Psychol.*, 1955, **19,** 445–448.

WOODROW, H., & LOWELL, Frances. Children's association frequency tables. *Psychol. Monogr.*, 1916, **22**(5, Whole No. 97).

(Received September 17, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

PRODUCTION OF ASSOCIATIVE SEQUENCES IN SCHIZOPHRENIA AND CHRONIC BRAIN SYNDROME[1, 2]

JOHN R. LESTER
Norwich State Hospital, Connecticut

ALTHOUGH Bleuler (1950, p. 32) was concerned with extreme time variations in the associative process as characteristic of pathological thinking, he had to acknowledge that, "We know almost nothing relative to the time relations of the schizophrenic associative process." Recently Arieti (1955) has reaffirmed our ignorance on this score.

In an effort to clarify Bleuler's notion of time differences in association among nosological groups, this study applied a relatively new technique to examine the temporal characteristics of production under conditions of restricted association and of recall. The associative productions of four groups (control, paranoid schizophrenic, hebephrenic schizophrenic, chronic brain syndrome epileptic) were considered within the framework of Hebbian Theory (1949).

Bousfield, Sedgewick, and Cohen (1954), employing normal subjects, developed a mathematical method to describe verbal production in terms of analytical scores. The unique feature of this method is that it permits an analysis of the efficiency of the production process independent of the content of associations. Where subjects are asked to name cities in the United States, for example, their responses occur readily at first, but the rate of production diminishes with time. Thus, within the limits set by the instructions, *S* has a supply of associates, and the rate at which this supply is depleted is proportional at any given moment to the number of associates remaining in reserve (Bousfield et al., 1954). When familiarity of the items to be recalled is high, the production of associates virtually ceases within 10 min., and the empirically obtained production curve can be expressed by an hyperbolic equation.[3]

It was recently suggested that the hebephrenic manages adjustment by a form of cortical activity termed "intrinsic organization" (Orgel, 1956, p. 26). Intrinsic organization involves: a heightened limen for sensory stimulation; a loss of purposeful, persistently directed activity; a reduction in memory; a loss of superordination (Hebb, 1949, p. 122). Since superordination is weak, the hebephrenic should manifest a marked deficit in efficiency of production. Deficit should, therefore, appear not only in the form of a reduced supply of associates, but also in a slow rate of depletion.

Similar predictions concerning the verbal productions of the epileptic should follow, for the behavioral correlates of intrinsic organization do seem to parallel the descriptive account of the epileptic's behavior (Orgel, 1956). However, one qualification is in order. It cannot be assumed that the psychopathology of all epileptics is as the simple consequence of cortical impairment. Functional factors, not unlike those in hebephrenia, can also be present. In some cases then, the efficiency of the epileptic's production may be doubly limited. Group data should show a reduced supply of associates, and the rate of depletion of even the limited supply should be slightly slower than that of the hebephrenics and markedly slower than that of the controls.

The behavioral correlates of intrinsic organization do not appear to be present in paranoid schizophrenics who have been reported to be

---

[1] This study is the condensed version of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Connecticut, June, 1957.

[2] The author is indebted to the members of his committee, Margaret M. Riggs, Weston A. Bousfield, James M. Sakoda, and Hermann O. Schmidt, for their encouragement and patient direction during the preparation of this study. In addition, the author wishes to thank Ronald H. Kettle, Superintendent, Norwich State Hospital, for his cooperative support; Jules D. Holzberg, Connecticut State Hospital, who graciously assisted in the selection of some of the subjects; and Sidney A. Orgel, Veterans Administration Hospital, Syracuse, for his helpful suggestions and for the loan of the apparatus.

[3] $n = c_2nt/l + cmt$: n is the total number of words produced in time, t; c is the asymptote approached by the function, rationally defined as the total number of items available for recall; m is a constant, the slope factor, the rate of depletion of the supply (Bousfield, Sedgwick, & Cohen, 1954, p. 116).

hyperalert (Rapaport, 1951; Berze, 1951), strongly motivated (Cameron, 1947), possessed of keen memory (Bleuler, 1950), or without a loss of superordination (Orgel, 1956). Rather, there are well-established phase sequences (thought changes) whose stability and resistance to modification testify to the domination of the delusional superordination. In Hebbian terms, well-developed phase sequences depend for their existence on continued but mild inconsistencies in sensory "feedback." Thus, the security operations of the paranoid seem to require constant, though distorted, interaction with others (Sullivan, 1953, p. 328). One may infer, therefore, that the association process of the paranoid schizophrenic is characterized by a minimum of disturbance in superordinate relationships, and a careful selectivity should be manifest in his associative sequences. Thus, though the supply of associates should be reduced, the paranoid should approach the control group in his rate of depletion.

As a working assumption, it was proposed that the rate of depletion reflects the amount of interference present during the selection of associates for production. The less the disruption of the phase cycles, the more firmly established is the superordinate structure; the less the interference, the faster is the rate of depletion and the more temporally efficient is the associative process, once allowance has been made for differences among the groups in the available supply of associates (gross output).

METHOD

Subjects

The *S*s were: 30 paranoid schizophrenics and 30 hebephrenic schizophrenics, chosen according to objective nosological criteria (Orgel, 1957); 30 chronic brain syndrome epileptics of unequivocal diagnoses,[4] and 30 controls with no known history of psychiatric disturbance. The four groups, each of 15 male and 15 female *S*s, were comparable for age, years of education, Stanford-Binet vocabulary IQ (Cureton, 1954), and length of hospitalization (institutionalized groups). The relevant data are presented in Table 1.

All 120 *S*s were examined individually. Patients were tested in a side room on the wards in which they resided. Control *S*s were examined in their homes. After *S* had entered the examining room, he was given a paper and pencil and asked to write his name and the sentence, "I hope to leave here very soon" (Babcock, 1930). When it was clear that he was able to comply with these instructions, he was presented with Task I.

TABLE 1

NORMATIVE DATE FOR EQUATED GROUPS

($N = 30$ in Each Group)

	Control	Paranoid	Hebephrenic	Epileptic
Chronological Age				
Mean	39.5	42.8	40.4	38.6
SD	6.9	5.2	7.4	6.4
Range	29–50	33–50	24–50	26–48
Years of Schooling				
Mean	10.3	10.0	9.7	9.1
SD	1.9	2.5	2.4	3.0
Range	7–14	7–19	6–16	6–16
Stanford-Binet Vocabulary IQ				
Mean	116.2	115.0	110.8	107.8
SD	14.6	14.3	12.8	12.7
Range	90–144	91–145	91–134	92–130
Months of Hospitalization				
Mean		135.9	142.2	119.3
SD		74.0	72.2	67.8
Range		34–320	45–302	33–352

Task I. Following Orgel's (1956) procedure, a single category (food) list of 40 two-syllable words was presented visually by memory drum. After three randomized presentations of the list, *S* was instructed to write down an many of the words as he could remember. *E* recorded the cumulative production at 15-sec. intervals for a period of 10 min.

Task II. After the termination of Task I, *E* said, "That was fine. Now I have something a little different for you to try." *E* presented *S* with a sheet of lined paper and gave the following instructions:

> When I say, "Go," I would like you to write down the names of all the colors that you can think of. You will have ten minutes to finish. After the first few minutes everyone finds it hard to think of any more colors. If this happens to you, you need not tell me about it and don't worry about it. Just sit back and try to remember more colors. Once we start, you are to write down as many names of colors as you can remember until I say, "Stop." Are there any questions?

When it was clear that the instructions were understood, *E* said, "Ready? Go." *E* recorded the cumulative production at each 15-sec. interval. Association was terminated after 10 min.

Pilot investigation had indicated that initial associations within the color category have a high degree of readiness, and the cumulative production curve virtually reached asymotote within ten minutes. This pattern of response was consistent for both well educated normals and psychotic *Ss*.

---

[4] Since the supply of female patients in this category was limited, it was necessary to select eight of this group from the Connecticut State Hospital. All other institutionalized subjects were residents of the Norwich State Hospital.

On termination of Task II, *E* administered Sections III and V of the Wechsler Memory Scale (Wechsler, 1945). Section III, Mental Control, was selected because the nature of the controlled association required is not unlike that demanded in Task II. Section V, Digit Span, was administered in standard fashion to obtain an independent index of immediate memory.

Classification of Responses

Bousfield (1954) in his method of analysis of the verbal productions of normal *Ss* classifies any written response as an instance of *recall* whether or not it actually corresponds to a word on the stimulus list. In the present study, a modification of this approach was introduced to take account of the intrusions which occur frequently in the productions of psychotics. Often these intrusions were inappropriate to the demands of the task: e. g., "saxophone" as part of the *recall* of a previously presented stimulus list of food words. Thus, for both Tasks I and II, analyses were performed at two levels:

Level 1: *Relaxed Standard*: To cover all written responses, including irrelevant intrusions, neologisms, perseverations, etc.

Level 2: *Rigid Standard*: To apply only to the first instance of each response which complies specifically with the task requirements.

Therefore, in Task I, Level 2, only the first occurrence of words which appeared on the stimulus list was credited. In Task II, Level 2, only the first occurrence of a word which actually appeared on either of two standard lists of color names (Kem-tone Color Charts;[5] Standard Color Card of America[6]) was acceptable. In all instances, where a response could not be found on the standard list of color names, *Webster's Collegiate Dictionary* (1937) was consulted as a final authority. If the dictionary definition of the questionable word included no reference to color, the response was classified as *irrelevant*.

RESULTS

The first step in analysis involved the computation of the cumulative totals of responses at the end of the first minute of production and at each successive 2-min. interval for each *S* on Tasks I and II, Level 1 and 2. The means and standard deviations of the cumulative totals of associates produced for the four nosological groups are presented in Tables 2 and 3. A summary of the results based on the average curves of the four nosological groups, Tasks I and II, Levels 1 and 2, appears in Table 4.

Total Product

For the analysis dealing with total product, both in the recall of a previously presented stimulus list (Task I) and under the condition of minimally restricted association (Task II), at the two levels of analyses, significant differences in productivity occurred among the groups ($p < .001$ in each of the four analyses of variance). Group means followed the predicted order—control, paranoid, hebephrenic, and epileptic in three of the four conditions. The one exception appeared in Task I, Level 2, where the epileptics produced more acceptable associates than the hebephrenics. Between-groups t tests revealed that the control group consistently differed from the patient groups in total production on Tasks I and II at both levels of analysis; among the patient groups, no other significant differences were obtained.

On the independent measures, Digit Span and Mental Control scores, the results followed the same pattern as the findings for Tasks I and II, i.e., the ordering of group means was in the same sequence. Only the controls differed significantly from the other groups; among the patient groups, no significant differences appeared.

Temporal Characteristics of the Association Process

Application of the equation for the hyperbola to each individual cumulative curve of production yielded score distributions for each group. Thus, for every *S*, values were obtained for c, m, and $t\frac{1}{2}$ (cf., Fn. 3) for the two tasks, each at two levels of analysis.

Bousfield (1954) defined m as an inverse measure of interference, i.e., other things equal, the greater the magnitude of m, the less frequently are associations interfering with each other during production, and the faster is the rate of depletion. There is the implication of "interference" or "friction" between competing associates which occurs as *S* successively selects and rejects verbal concepts for appropriateness of response. Selection consumes time; thus, with supply constant, the more smoothly the selection proceeds, the greater is the magnitude of m, and the faster is the rate of depletion.

Differences among the groups in rate of depletion can be meaningfully evaluated only

---

[5] Kem-tone Commercial Color Charts. Detroit: Rogers Co., 1956.

[6] The Textile Color Card Association of the United States, Inc. *Standard color card of America* (9th ed.) New York: 1941.

TABLE 2

MEANS AND *SD* OF TOTALS OF WORDS PRODUCED AS A FUNCTION OF TIME SPENT IN PRODUCTION FOR THE FOUR NOSOLOGICAL GROUPS: TASK I, LEVEL 1 (n) AND LEVEL 2 ($n - i$)
(N = 30 in Each Group)

Group		Time (in Min.) After Starting Signal											
		1		2		4		6		8		10	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Level 1 (n)	Control	11.57	2.78	15.33	3.66	18.07	4.26	19.70	5.09	21.13	5.15	21.70	5.32
	Paranoid	8.03	2.50	10.03	3.51	12.60	4.30	14.00	4.77	15.23	4.80	15.87	5.10
	Hebephrenic	6.53	3.41	8.80	3.92	11.10	4.90	12.53	5.32	13.90	6.19	15.47	7.38
	Epileptic	5.53	2.35	8.67	3.41	11.17	4.15	12.63	4.55	13.73	5.01	14.53	5.38
Level 2 ($n - i$)	Control	11.53	2.81	15.23	3.74	17.70	4.16	19.57	5.03	20.40	5.11	20.93	5.26
	Paranoid	7.63	2.35	9.60	3.43	12.10	4.07	12.97	4.76	13.57	4.88	14.03	5.13
	Hebephrenic	6.20	3.54	7.97	4.12	9.57	4.75	10.37	5.03	11.03	5.25	11.47	5.46
	Epileptic	5.30	2.49	8.27	3.55	10.60	4.22	11.80	4.59	12.60	4.89	13.03	5.09

TABLE 3

MEANS AND *SD* OF TOTALS OF WORDS PRODUCED AS A FUNCTION OF TIME SPENT IN PRODUCTION FOR THE FOUR NOSOLOGICAL GROUPS: TASK II, LEVEL 1 (n) AND LEVEL 2 ($n - i$)
(N = 30 in Each Group)

Group		Time (in Min.) After Starting Signal											
		1		2		4		6		8		10	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Level 1 (n)	Control	11.80	2.39	16.30	3.53	21.37	5.56	25.37	7.13	28.40	9.13	30.93	10.44
	Paranoid	9.53	1.71	11.83	2.42	15.30	3.73	17.17	4.37	18.33	5.13	19.33	5.74
	Hebephrenic	7.73	3.09	10.43	3.80	13.63	5.48	15.33	6.70	16.40	7.34	17.77	8.26
	Epileptic	7.27	3.09	10.00	3.64	12.60	3.76	14.27	4.41	16.17	5.25	17.43	5.85
Level 2 ($n - i$)	Control	11.77	2.33	16.23	3.36	21.07	5.18	24.93	6.58	27.93	8.66	30.40	9.96
	Paranoid	9.50	1.74	11.80	2.41	15.03	3.32	16.80	3.94	17.83	4.42	18.57	4.74
	Hebephrenic	7.70	2.99	10.17	3.66	13.17	5.25	14.73	6.47	15.63	7.05	16.67	7.90
	Epileptic	7.23	3.08	9.63	3.28	11.90	3.56	13.00	3.85	14.33	4.42	15.03	5.00

TABLE 4

DERIVED CONSTANTS AND HALF-LIVES FOR RATIONAL EQUATION FITTED TO GROUP DATA FOR FOUR NOSOLOGICAL GROUPS: TASKS I AND II, LEVELS 1 (n) AND 2 ($n - i$)

		c		m		$t_{1/2}$	
		n	$n - i$	n	$n - i$	n	$n - i$
Task I	Control	24.18	23.27	.0340	.0381	1.22	1.13
	Paranoid	18.18	15.55	.0346	.0561	1.59	1.15
	Hebephrenic	18.01	12.69	.0255	.0659	2.18	1.20
	Epileptic	17.48	15.38	.0265	.0362	2.16	1.80
Task II	Control	38.81	37.86	.0090	.0095	2.87	2.78
	Paranoid	22.28	21.28	.0267	.0306	1.68	1.54
	Hebephrenic	20.72	19.25	.0248	.0298	1.95	1.74
	Epileptic	20.72	17.08	.0214	.0370	2.26	1.58

after the influence of group differences in available supply have been controlled. For Tasks I and II, Levels 1 and 2, analyses of covariance were performed with c as the independent variable and m as the dependent variable. The results of the four analyses of covariance were as follows: Task I, Level 1, $F = 11.78, p < .001$; Task I, Level 2, $F = 8.30$,

TABLE 5

UNADJUSTED MEAN INDICES OF INTERFERENCE AND MEAN INDICES OF INTERFERENCE ADJUSTED FOR DIFFERENCES IN THEORETICAL SUPPLY OF ASSOCIATES: TASKS I AND II, LEVELS 1 AND 2

Level	Task I				Task II			
	1		2		1		2	
Mean	Adj.	Unadj.	Adj.	Unadj.	Adj.	Unadj.	Adj.	Unadj.
Control	.0605	.0356	.0861	.0422	.0323	.0124	.0387	.0130
Paranoid	.0355	.0363	.0599	.0703	.0273	.0326	.0329	.0352
Hebephrenic	.0261	.0325	.0452	.0689	.0223	.0351	.0274	.0421
Epileptic	.0237	.0289	.0329	.0374	.0177	.0244	.0260	.0450

$p < .001$; Task II, Level 1, $F = 3.82$, $p < .05$; Task II, Level 2, $F = 1.38$, *ns*. The adjusted and unadjusted mean indices of interference appear in Table 5. For Task I, Levels 1 and 2, and Task II, Level 1, six analyses of covariance compare the groups two at a time. The *F*s obtained by these analyses are presented in Table 6.

These findings indicate that differentiation between groups, with regard to the amount of interference manifest, increased directly with the number of the restrictions placed on association. Under the condition approaching free association, if inappropriate responses were ignored (Task II, Level 2), no differences were evident in interference. As increased restrictions were imposed on the association process, a continuum of deficit in the ease of synchronization of associates appeared. Maximum differentiation among groups was achieved in Task I, Level 1. The continuum which emerged as an important result of the four analyses is presented in Table 7.

Efficiency of Production

Efficiency ($t\frac{1}{2}$), as here employed, is not a measure of simple rate of production of associates. Rather, it is based on the relationship between the supply of items available for production and the amount of interference present during production, i.e., $t\frac{1}{2} = 1/cm$. Since c and n are highly correlated, n was chosen as the independent variable for covariance analysis, not because it is more effective than c in exerting control on $t\frac{1}{2}$, but because both n and $t\frac{1}{2}$ are traditionally empirical measures of production and efficiency respectively (Bousfield, Sedgewick, Cohen, 1954; Johnson 1955).

TABLE 6

*F*S OBTAINED IN TESTS OF DIFFERENCES BETWEEN ADJUSTED MEAN INTERFERENCE INDICES FOR SIX SETS OF COMPARISONS BETWEEN NOSOLOGICAL GROUPS: TASK I, LEVELS 1 AND 2; TASK II. LEVEL 1

Groups	Task I		Task II
	Level 1	Level 2	Level 1
Control vs. Paranoid	13.45[c]	3.50	3.25
Control vs. Hebephrenic	26.16[c]	8.43[b]	2.08
Control vs. Epileptic	31.38[c]	19.76[c]	10.36[b]
Paranoid vs. Hebephrenic	3.64	2.18	1.15
Paranoid vs. Epileptic	5.12[a]	7.84[b]	8.22[b]
Hebephrenic vs. Epileptic	.28	9.32[b]	3.99

[a] $p < .05$.
[b] $p < .01$.
[c] $p < .001$.

The results of the four analyses of covariance ($n = X$; $t\frac{1}{2} = Y$) are: Task I, Level 1, $F = 7.45$, $p < .001$; Task I, Level 2, $F = 4.33$, $p < .01$; Task II, Level 1, $F = 1.93$, *ns*; Task II, Level 2, $F = 0.71$, *ns*. The adjusted and unadjusted mean indices of efficiency appear in Table 8. For Task I, Levels 1 and 2, six separate analyses compare the groups two at a time. *F*s obtained by these analyses are presented in Table 9.

For Task I, Level 1, the controls proved to be more efficient than all patient groups, after the influence of sheer output of associates was removed. At this level of analysis, the paranoids surpassed the epileptics in efficiency. The hebephrenic group lay between the paranoid and epileptic groups, but differed significantly from neither.

In Task I, Level 2, after intrusions were removed, the control vs. epileptic and paranoid vs. epileptic comparisons revealed the only significant differences. The fact that the hebephrenic and paranoid groups no longer differed from the control group at this level of analysis,

TABLE 7

PATIENT GROUPS WHICH MANIFEST A SIGNIFICANTLY GREATER AMOUNT OF INTERFERENCE WHEN COMPARED WITH THE CONTROL GROUP AS A FUNCTION OF THE TASK AND LEVEL OF ANALYSIS

Task	Level	Groups
I	1	Paranoid, Hebephrenic, Epileptic
	2	Hebephrenic, Epileptic
II	1	Epileptic
	2	None

TABLE 8

UNADJUSTED MEAN HALF-LIVES AND MEAN HALF-LIVES ADJUSTED FOR DIFFERENCES IN PRODUCTION: TASKS I AND II, LEVELS 1 AND 2

	Task I				Task II			
	Level 1		Level 2		Level 1		Level 2	
	Adj.	Unadj.	Adj.	Unadj.	Adj.	Unadj.	Adj.	Unadj.
Control	.92	1.19	.91	1.08	1.73	2.42	1.44	2.29
Paranoid	1.50	1.37	.98	.96	1.50	1.44	1.38	1.36
Hebephrenic	1.99	1.84	1.34	1.22	1.78	1.51	1.67	1.37
Epileptic	2.31	2.03	1.86	1.77	2.33	2.04	1.72	1.34

TABLE 9

*F*s OBTAINED IN TESTS OF DIFFERENCES BETWEEN ADJUSTED MEAN HALF-LIVES FOR SIX SETS OF COMPARISONS BETWEEN FOUR NOSOLOGICAL GROUPS: TASK I, LEVELS 1 AND 2

Groups	Level 1	Level 2
Control vs. Paranoid	5.20^{a}	1.54
Control vs. Hebephrenic	4.65^{a}	1.36
Control vs. Epileptic	16.30^{c}	8.22^{b}
Paranoid vs. Hebephrenic	1.54	1.94
Paranoid vs. Epileptic	4.16^{a}	8.07^{b}
Hebephrenic vs. Epileptic	.46	2.76

[a] $p < .05$.
[b] $p < .01$.
[c] $p < .001$.

as was the case at Level 1, suggests that it is the presence of intrusions which accounts for a significant reduction in efficiency in both the hebephrenic and paranoid groups at Level 1.

Because $t\frac{1}{2}$ depends directly on m for its power of differentiation, $t\frac{1}{2}$ is best considered as a summary score. The value of $t\frac{1}{2}$ as a measure lies in its ability to express a complex interaction between interference and the theoretical supply in units of time. However, unadjusted half-lives tend to be misleading (cf., ordering of unadjusted mean half-lives vs. order of adjusted mean half-lives for all groups, Table 8). Thus, it is only after adjustment for differences in output that half-lives provide meaningful comparisons *when applied to groups grossly differing in volume of production.*

TABLE 10

NUMBER OF SUBJECTS IN EACH GROUP ($N = 30$) WHO PRODUCE ONE OR MORE IRRELEVANT INTRUSIONS (TASKS I AND II)

	Irrelevant Tasks I	Irrelevant Tasks II	Categorical Task I
Control	0	6	6
Paranoid	3	4	16
Hebephrenic	10	10	15
Epileptic	8	10	10

Intrusions

Separate chi square analyses were performed for the number of *S*s in each group having no irrelevant intrusions vs. the number having one or more irrelevant intrusions (Tasks I and II), and for the number of *S*s giving no categorical intrusions vs. the number giving one or more categorical intrusions (food words not on the list) in Task I. The data for the production of intrusions in the three situations is presented in Table 10.

The over-all chi square for the number of *S*s who gave one or more irrelevant intrusions were: Task I, 14.87; Task II, 13.34 (which in both cases for 3 *df* are significant beyond the .01 level). In Tasks I and II chi square comparisons between the groups, two at a time (Yates correction 1 *df*), yielded the values presented in Table 11.

The over-all chi square for the number of *S*s who produced one or more categorical intrusions in Task I was 9.07, 3 *df*, $p < .05$. In chi square comparisons between groups, two at a time (Yates correction 1 *df*), significant differences were as follows: control vs. paranoid, $p < .02$; and control vs. hebephrenic, $p < .05$.

In terms of percentage of the total product in Task I, the hebephrenic and paranoid groups manifested a stronger tendency to give categorical intrusions than appeared in the control and epileptic groups. Perseverations made up a

TABLE 11
p VALUES FOR CHI-SQUARE COMPARISONS BETWEEN GROUPS ON NUMBER OF SUBJECTS WHO GIVE ONE OR MORE IRRELEVANT INTRUSIONS (TASKS I AND II)

		Control	Paranoid	Hebephrenic	Epileptic	
Task I	Epileptic	.05	*ns*	*ns*		Task II
	Hebephrenic	.01	.10		*ns*	
	Paranoid	*ns*		*ns*	.01	
	Control		*ns*	*ns*	.02	

relatively large proportion of the intrusions produced by epileptics, particularly in Task II.

For all groups in both tasks, the absolute number of intrusions did not increase appreciably with time from the starting signal. However, there was a general tendency for all groups to give a higher percentage of intrusions as the output of appropriate associates decreased over time. On Task I, this effect in the hebephrenic group was quite marked; in the 8–10 min. interval almost 75% of the responses produced were words other than those which appeared on the original stimulus list. In Task II, it was the epileptics who showed the strongest tendency to give an increasingly higher percentage of irrelevant and perseverative intrusions, as time from the starting signal increased.

The findings with regard to intrusions can be related to the earlier results on group differences in interference and efficiency. Groups which introduce the greater number of intrusions toward the end of the production period also show a greater relative improvement (decrease in interference, increase in efficiency) when intrusions are removed, because, with intrusions eliminated the production curves of these groups tend to reach asymptote more quickly.

Nevertheless, the results of the analyses did reveal significant differences among the groups even after the intrusions were removed. Thus, the intrusions do not by themselves account for all differences among the groups on the measure of interference and efficiency.

DISCUSSION

While the temporal characteristics of the production process defined the general area of interest in this study, it should be apparent that it is the interference constant, m, which stands as the most significant index. Two variables, interference (m) and theoretically available supply (c), were derived from the cumulative production curve as separate aspects of the same data, and they shared considerable common variance (of the order $r = .80$). Yet when the variance in m common to c was removed, the residual variance in m as a measure of interference provided an excellent basis for differentiation among the groups.

The controls were found to differ from the patient groups with respect to gross output of words under the four conditions of measurement, confirming earlier studies reporting a greater total output for normals as compared to patient groups. However, where the focus shifted to group differences in rate of response (after allowance was made for difference in supply), differentiation between the groups clearly improved.

Since it was the hebephrenic and epileptic groups who manifest the behavioral correlates of intrinsic organization, as opposed to the paranoids who do not, the prediction was made that the hebephrenics and epileptics would reveal more evidence of disturbance in superordination than the paranoids or controls. The results, in general, tend to support this position derived from Hebb. It was the epileptics and hebephrenics who showed considerable interference in the selection of associates, which appeared as a slowing in the flow of the available responses. On the grounds that paranoids are characterized by a "lasting modification of thought," which, however, does not imply impaired superordination, it was predicted that the paranoids would approach the controls with respect to efficiency of superordination. The paranoids did approach the controls, and both these groups revealed much less interference in their associative sequences than the hebephrenics and epileptics.

In terms of the amount of interference, once allowance had been made for differences in the available supply, the predicted ordering from most to least (epileptic, hebephrenic, paranoid, control) was consistently confirmed for the recall situation (Task I) and in the more general condition of restricted association (Task II), whether the appropriateness of the responses (Level 1 vs. 2) was considered or not.

In this experiment, *S*, in order to perform efficiently, must adopt one set at a time, i.e., (Task I) name only foods on the stimulus list or (Task II) name only colors. Thus, where intrusions occur, there is behavioral evidence that interfering systems have invaded the leading system (Angyal, 1941, p. 309). The patient groups, particularly the hebephrenics and epileptics, seemed unable to resist the invasion of competing systems (quite possibly the systems associated with the emotional complexes, as Jung [White, 1944] suggested). On the other hand, intrusions seem to be but symptoms of the underlying weakening of integration, and they project above the surface like the top of an iceberg. More significant for an understanding of deviant thought is the fact that these interfering systems, of which intrusions are only obvious signals, "continue to pound at the leading systems" (Angyal, 1941) and create resistance or friction which slows down the thought process and appears in this experiment as a decrease in the magnitude of the slope constant, m (a slowing in the rate of depletion). The slowing effect was reduced for all groups when intrusions were ignored; but even then significant slowing was still detectable in the pathological groups, and differentiation among some of the groups was still possible with this measure.

The half-life, as a measure of efficiency, proved to be effective in ordering the groups, once an allowance was made for differences in volume of output. Babcock (1930, p. 42) commented on the fact that individuals who have a limited supply, such as the cautious paranoids, may be efficient in terms of the speed of their associations. Babcock's position is supported by present findings, particularly in Task II, where the paranoids produced significantly fewer responses than the control but still proved to be as efficient. Consistent with Babcock's position, Bousfield, Sedgewick, & Cohen (1954) cautioned that efficiency must depend not only on the size of the supply but also on the amount of interference.

Heretofore, the problem of precise and appropriate measurement of speed of verbal response has been difficult, since only gross measures of total production were employed. As demonstrated in this experiment, the analytical technique developed by Bousfield, Sedgewick, & Cohen (1954) contributes a new dimension to such analyses, taking into account temporal aspects of the process of production as well as the end product. In the last analysis, these results are in the nature of empirical observations which do not contradict previous findings in the general area of associative production. Rather, they serve to illuminate the interpretation of other empirical findings by increasing our understanding of certain previously unexplored deviations in the thought processes of the groups concerned.

SUMMARY

This study was concerned with timing differences in the associative processes of schizophrenics and epileptics. A recall situation and a more general condition of restricted association were employed to elicit verbal sequences which were analyzed in terms of efficiency of the production.

The controls were found to differ from the patient groups with respect to gross output of words, but where the analysis took account of the temporal aspects of the process of production, differentiation between groups clearly improved. As increased restrictions were imposed on the association process, disturbances in the thinking of the pathological groups became more apparent.

The results suggest that inappropriate responses (intrusions), which appear in the productions of psychotic individuals, are but symptomatic of a more basic disturbance in thought processes. The interference measure employed in this study seems to provide a more direct index of the degree of thought disorganization.

The assumption that interference, resulting from the pressure of extraneous associations, hampers associative production by slowing the rate of responding is discussed within the framework of Hebbian theory and in relation to other theoretical formulations.

REFERENCES

ANGYAL, A. *Foundations for a science of personality.* New York: Commonwealth Fund, 1941.

ARIETI, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.

BABCOCK, HARRIET. An experiment in the measurement of mental deterioration. *Arch. Psychol., N. Y.,* 1930, **18,** No. 117.

BERZE, J. Das primaersymptom der paranoia. In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: International Univer. Press, 1950.

BOUSFIELD, W. A., & SEDGEWICK, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.

BOUSFIELD, W. A., SEDGEWICK, C. H. W., & COHEN, B. H. Certain temporal characteristics of the recall of verbal associates. *Amer. J. Psychol.*, 1954, **67**, 111–118.

CAMERON, N. *The psychology of behavior disorders.* New York: Houghton Mifflin Co., 1947.

CURETON, E. E. Mental age equivalents for the Revised Stanford-Binet vocabulary test. *J. consult. Psychol.*, 1954, **18**, 381–383.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

ORGEL, S. A. Clustering of verbal associates in schizophrenia and the chronic brain syndrome. Unpublished doctoral dissertation, Univer. of Connecticut, 1956.

ORGEL, S. A. Differential classification of hebephrenic and paranoid schizophrenics from case material. *J. clin. Psychol.*, 1957, **13**, No. 2, 159–161.

RAPAPORT, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.

SULLIVAN, H. S. *The interpersonal theory of psychiatry.* New York: W. W. Norton, 1953.

Webster's collegiate dictionary. (5th Ed., Abridged) Springfield, Mass.: G. & C. Merriam, 1937.

WECHSLER, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

WHITE, R. W. Imaginative productions. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944.

(Received September 12, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

VERBAL AND GSR RESPONSES IN EXPERIMENTAL INTERVIEWS AS A FUNCTION OF THREE DEGREES OF "THERAPIST" COMMUNICATION

BARCLAY MARTIN, RICHARD M. LUNDY, AND MARK H. LEWIN[1]
University of Wisconsin

THE present study deals with the tendency of a subject in an experimental psychotherapy interview to approach or avoid coming to grips with therapeutically significant material as a function of the amount of communication he receives from the therapist. The process of psychotherapy is formulated in terms of an approach-avoidance model similar to the conceptions of Miller (1944) and Murray and Berkun (1955), in which it is conceived as a situation where the client alternately approaches significant material and retreats from it. No single motive is postulated to account for the approach behavior, but rather several motives that vary in nature from client to client. Thus, the long range goal of relief from psychological distress may motivate a person to endure the momentary increases in discomfort usually associated with approach behavior. There are also more immediate motives, such as a desire to please and gain approval from the therapist, a desire to obtain advice and reassurance, etc. Avoidance behavior, on the other hand, is motivated by the anxiety, embarrassment, or (more generally) the negative affect associated with approaching therapeutically significant material. It is postulated that, as a client alternately approaches and retreats from such material and experiences the resultant negative affect in a therapeutic atmosphere of noncritical acceptance, understanding, etc., the negative affect associated with these sensitive areas tends to extinguish. It is also assumed that this extinction of negative affect generalizes to some extent beyond the therapy session, and that, accordingly, this process is in part responsible for what is therapeutic about psychotherapy.

A question that naturally follows from these considerations is, "What are some of the factors which facilitate or impede approach behavior?" The factor that was introduced experimentally into the present study was the amount of communication or feedback that the *S* received from the therapist. Three groups were employed that were thought to represent points along this dimension. In one group, the *S* talked to a tape recorder with no therapist present, and accordingly received virtually no communication from the therapist. In a second group, the therapist was present but responded by nonverbal means only. In the third group, the therapist responded in an essentially client-centered fashion, thus providing a maximum of communicative feedback compared with the other two treatment groups.

It was thought that an *S* in Group 1 who received no communication from the therapist when he engaged in approach behavior would soon extinguish and show an increasing tendency to avoid therapeutically significant material. This was predicted because there would be no positive reinforcement from the nonpresent therapist for approach attempts, nor would the *S* have any way of knowing to what extent the nonpresent therapist might be implicitly responding in a critical, judgmental fashion. The nonverbal therapist condition, by providing some communication, was expected to result in more approach behavior than in Group 1, but less than in the "regular" therapy condition of Group 3.

These expected relationships were to be analyzed in terms of approach and avoidance behavior trends over the five experimental sessions. Thus, the research can be viewed in a more restricted sense as a study of the reinforcing effects of three degrees of experimenter communication upon affectively toned verbilizations in the *S*s. In addition, changes in tension or anxiety level, both within and between sessions, were assessed by continuous GSR recordings. Since various alternative predictions with respect to increasing or decreasing tension levels among the treatment groups seemed reasonable, no specific hypothesis was proposed with respect to this part of the study.

---

[1] The authors wish to thank Irene Waskow, Vilma Ginsberg, and Jack Hokanson for their helpful assistance on this study, and the University of Wisconsin Graduate School for providing financial support for research leave for one of the authors.

METHOD

Subjects

The Forced Choice Manifest Anxiety Scale (Heineman, 1953) was given to the 240 students in a course on the Psychology of Adjustment. At the same time, the students were asked to indicate on their IBM answer sheets if they would be interested in participating in a brief psychotherapy experience as part of a research project. From the students who volunteered, (about 80% of the class) the 44 who scored highest on the anxiety scale were contacted for 15-min. preliminary interviews. At the end of each interview, the interviewer rated *S* on a 5-point scale of severity of disturbance and a 5-point scale of defensiveness. No *S*s were eliminated because of the possible eminence of a psychotic break; one *S* was eliminated because she was already on the waiting list at another clinic. None of the *S*s had experienced psychotherapy within the last two years.

The first step was to select the 27 *S*s who were relatively high on disturbance and low on defensiveness. Next, *S*s were divided into groups of three, with the 3 *S*s in each group matched as well as possible on their disturbance and defensiveness ratings. Within each triad, they were then randomly assigned to one of the three major treatment groups. For each treatment group, 5 *S*s were assigned to therapist A, 3 to therapist B, and 1 to therapist C. Ten *S*s were male, 17 female.

The Treatment Procedures

Tape recordings and continuous GSR recordings were obtained for all *S*s in all sessions, including the preliminary session. The GSR apparatus consisted of an ac circuit designed by Grant (1946) which fed into a G.E. Photoelectric Recorder. Dry silver electrodes were attached to the palmar surfaces of two of *S*'s left hand fingers, which rested comfortably on a low table beside the *S*. The GSR and tape recordings were synchronized by a timer which operated a marker pen on the GSR record and simultaneously made an audible click in the tape recorder every 37 sec.[2]

*S*s were seen for five 30-min. sessions one week apart. At the beginning of the first experimental session, *S*s were given instructions that varied with the treatment condition as described below:

> *Tape Recorder Group*. We are asking the subjects in different parts of the study to do different things. As you know, we are interested in studying various factors which affect the process of psychotherapy, and in one group, the group you are in, we want to study the effects of not having a therapist present in the room at all. We would like for you to imagine that you are a client in psychotherapy, but instead of talking to a therapist we want you to talk into this tape recorder. This may seem a little strange, but try to talk about the things that you would expect to in a regular therapy session. People ordinarily talk about things that are emotionally important to them and how they get along with the important people in their lives. But in general, feel free to talk about whatever comes to mind. There may be times when nothing comes to mind, and that's O.K. too. You can just wait until something does, and then talk about that. In general, though, try to keep talking as much as you can. After each session I will listen to the tape recording of your session. I will not talk with you about the session, though. I will just start you going each time, but will not talk with you about what you say on the tapes, although I will be listening to them after each session.

Nonverbal Group. These instructions were the same except for the following modification:

> . . . We want to study the effects of having you talk to a therapist who remains silent. So that, as the therapist in this situation, I will be listening carefully to what you have to say, but will respond by nonverbal means only. In other words, I will indicate that I understand what you are saying by nodding my head, or saying 'uh-huh', etc., but I will not be able to say anything out loud.

Regular Group. These instructions were similar to the above except that the *S* was told that the therapist would be responding as he would in ordinary psychotherapy.

Although outcome assessment was not the primary focus of the study, a forced-choice adjective check list, hereafter referred to as the Feeling Inventory, was obtained on all *S*s after the preliminary interview and again at the end of the fifth session. This inventory consisted of 41 sets of three adjectives each with one adjective in each set representing a tension or anxiety dimension. *S* is asked to indicate which adjective is most and which is least like the way he feels. Some evidence for the sensitivity of this inventory to changes in anxiety level had been previously established (Martin, 1959).

*S*s were instructed to answer the Feeling Inventory before the experimental sessions in terms of how they had been feeling during the past month. They were instructed at the end of the sessions to answer in terms of how they had been feeling during the last several days. A fourth group of nine *S*s who were matched roughly on initial score on this inventory with the other groups, but who did not attend any experimental sessions, were retested after the same amount of time had elapsed.

The Content Ratings of the Interviews

The following rating scheme was developed after considerable trial-and-error practice on other tapes. Each 74-sec. interval of an interview was rated individually for two major categories of behavior: Approach and Avoidance. If approach behavior occurred during an interval, a check was placed under the Approach column on the rating sheet; if avoidance behavior also occurred during the same interval, a check was placed in the Avoidance column.

---

[2] The 37-sec. interval has no particular significance. By mistake, the timer was set for this interval rather than 30 sec., and had to be continued at this interval throughout the study.

In addition to this rather arbitrary time interval and the dichotomous scoring of approach-avoidance behavior, an additional rating of the extent of approach behavior based on more meaningful units of behavior was made. To this end, a four-point scale was developed for assessing degree of approach to therapeutically meaningful material. Raters were required to classify the approach behaviors checked under the Approach Column according to meaningful themes and to make a rating on the four-point scale for each of these "meaning units."

Some examples of the kinds of behavior representing avoidance and approach behavior at the four levels of the rating scale are given below. An attempt was made to make the four-point scale appropriate to the range approach behaviors expected in this particular study and sample, and it is likely that some modifications would be necessary before applying this scheme to the process of longer, regular psychotherapy.

Avoidance Behavior. General discussion of course work, campus activities, and other ordinary school-related topics without relation to thir own personal-emotional conflicts; abstract philosophical or psychological discussion; description of other people and their personalities for at least one 37 sec. interval without relating this to their own problems; description of only positive or success experiences for at least one 37 sec. interval; any blanket and categorical denial of negative feelings with respect to a given person or situation.

Approach behavior, Level 1. In general, indications of approach were inferred both from signs of increasing negative affect resulting from the approach, and signs of verbal approach based upon the raters' preconceptions about therapeutically significant material. The following are examples of Level 1 approach: (*a*) talks in generalities about self, (*b*) describes interpersonal relationships including those with parents; (*c*) describes historical background material; (*d*) no feeling or strong personal involvement is expressed; (*e*) may specify negative feelings, but qualifies with instances of positive or success experience and does not indicate strong conflicts.

Approach behavior, Level 2. Describes clear cut psychological "symptom" or problem area, for example: (*a*) impotency, nervous vomiting, or specific situations in which handicapped by shyness, anxiety, hostility, etc.; (*b*) does not necessarily experience much feeling while describing symptom; (*c*) employs less qualification and less emphasis on positive or success experiences than for Level 1; (*d*) problem may be externalized to large extent.

Approach behavior, Level 3. Considers interpersonal aspects of symptoms or problem with some specificity accompanied by feeling; this feeling may relate to emotional reactions associated with the particular conflict area or may relate to feelings of discomfort and anxiety about going into the area; highly charged expression of feeling without much content exploration also scored here.

Approach behavior, Level 4. Shows strong, unequivocal expression and experiencing of feeling associated with approach to conflict area with considerable specificity. Level 4 represents an extension and intensification of Level 3 behavior.

The 27 *S*s were assigned in groups of nine to each of three judges for rating purposes. Two of these judges were therapists in the research and were assigned *S*s they had *not* seen in the interviews. Whenever a judge rated an interview, another person set the tape recorder up for him so that he would not know which interview in the sequence was being rated. If there were initial comments in the interview that would provide a clue as to the sequence of the interview, the tape was set so that the rating began after this point. The first, second, fourth, and fifth interviews were rated with all those for a given *S* done at one time. It was not possible to keep the raters blind with respect to the treatment condition, but they were blind with respect to sequence.

Reliability estimates of the scores obtained from these ratings were obtained as follows. Judges A and B jointly, but independently, rated interviews for five *S*s selected randomly; judges B and C, likewise, rated five other *S*s. The three basic scores used were: (*a*) Meaning Unit Rating, based upon the products of meaning unit rating (1–4), times the number of the 74-sec. intervals involved in the meaning unit and summed for all meaning units for a given interview; (*b*) Approach Behavior, the number of intervals during which approach behavior was checked in an interview; (*c*) Avoidance Behavior, the number of intervals during which avoidance behavior was checked in an interview. A fourth score, change in Meaning Unit rating from the first half to the last half of each interview, was also obtained. As an estimate of reliability, the three basic scores were correlated for the two judges for *each S* over the four interviews, because it was changes within *S*s with which the study was concerned. Thus, for judges A and B, there would be a correlation based on four pairs of scores for each of the five *S*s for each measure. These correlations were averaged, using Fisher's z transformation; the averages ranged from .80 to .94.

RESULTS

Curves for the average Meaning Unit ratings and intra-interview change in Meaning Unit ratings as a function of the four rated interviews are shown in Fig. 1. These curves depict

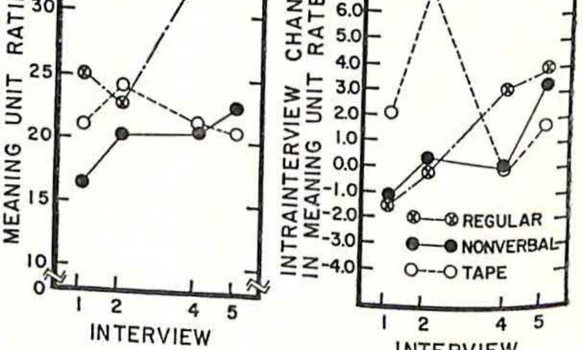


FIG. 1. Meaning unit ratings as a function of successive interviews.

TABLE 1

MEANS AND SIGMAS FOR THE PRE- AND POST-ANXIETY SCORES ON THE FEELING INVENTORY

Group	Preexperiment		Postexperiment	
	Mean	Sigma	Mean	Sigma
Regular	19.6	5.4	14.1	4.2
Nonverbal	16.3	7.2	11.2	6.1
Tape	16.6	5.9	14.3	3.8
Control	16.7	3.3	17.6	3.4

the Regular group as showing the greatest increase in both average Meaning Unit ratings and in intra-interview change in Meaning Unit ratings as the interviews progressed. The Tape group showed a slight tendency to decrease in both these measures, and the Nonverbal group revealed an intermediate amount of change.

The results for the Approach and Avoidance ratings showed similar between-group trends, but because statistical analysis showed no significant between-group effects, the curves for these variables are not presented.

For purposes of statistical analysis of the trends shown in Fig. 1, the ratings for interviews one and two and those for interviews four and five were averaged. Differences between these two sets of averages were then computed, and an analysis of variance was performed to see if the treatment groups differed significantly in their average change scores. An orthogonal polynomial trend test was not performed on these data because the assumption of equal intervals was not met, interview three not having been rated. The between-group changes for the average Meaning Unit ratings did not reach statistical significance. The between-group changes in intra-interview change in Meaning Unit ratings, however, were significantly different ($p <$ .05). A Multiple Range Test (Duncan, 1955) indicated that the Regular and Tape groups were significantly different from each other, but neither differed significantly from the nonverbal group. Thus, if one considers the change from the first half to the last half of an interview in approach to therapeutically significant material, one finds that the Regular group showed a progressively increasing tendency to approach further than did the Tape Group.

The means and standard deviations for the pre- and postexperiment results on the Feeling Inventory are shown in Table 1. Although the initial mean score for the Regular group is higher than the means for the other groups, it was not significantly so as tested by the Multiple Range Test. These Feeling Inventory scores were converted into pre- and postexperiment difference scores, and the Multiple Range Test was then performed. The Nonverbal and Regular groups both showed significantly greater decreases in reported anxiety than did the Control group, whereas the Tape group was not significantly different from the Control group on this score.

The skin conductance recordings were scored both in terms of general level of conductance and in terms of reactivity. Median log conductance scores were obtained for all five interviews for each *S*. A change in log conductance score was also obtained for each interview by subtracting the median log conductance score for the first half of the interview, from the last half of the interview. A reactivity score was obtained for each interview by marking off four equally spaced 4-min. intervals in each interview, obtaining the average of the three largest GSR deflections within each log conductance units, and then averaging these four values for an over-all estimate of reactivity. A change in reactivity score was obtained by subtracting the average reactivity score in the first 4-min. period from that in the last 4-min. period. Trends of these four scores as a func-

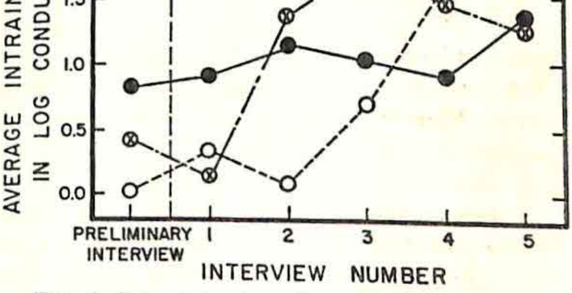


FIG. 2. Intra-interview change in log conductance as a function of successive interviews.

tion of the five sessions were tested for statistical significance by orthogonal polynomial trend tests (Grant, 1956). A significant between-groups trend was found only for the intra-interview change in conductance scores, where a significant between-groups quadratic component emerged. Inspection of the curves in Fig. 2 shows that the Regular group showed an initial rise and then a fall in the variable, whereas the Tape group showed an initial fall and then a sharp rise in this variable, with the Nonverbal group lying in between.

DISCUSSION

It is perhaps not altogether appropriate to refer to these experimental sessions, even those of the Regular group, as psychotherapy. The conditions and *S*s were rather special and not necessarily representative of psychotherapy or psychotherapy clients. The more restricted problem of studying factors that affect the extent to which a person openly discusses emotionally important aspects of his life is, however, certainly one that has direct implications for the more complex therapeutic process.

The intra-interview change in Meaning Unit ratings showed significant between-group changes from the first two to the last two interviews. Stated otherwise, the Regular group showed a marked increase in approach to emotionally important material from the first to the last half of an interview as the five interviews progressed, whereas the Tape group, although rather erratic, showed a slight tendency to decrease on this measure over the interviews. Perhaps these results reflect the reinforcing effect of an attentive, responding therapist who communicates not only positive reinforcement for approach behavior, but also by his accepting, noncritical attitude aids in the extinction of the negative affect associated with the approach behavior. For this short series of experimental interviews, this "learning" appears more pronounced within interviews than when between-interview means are compared. Since it is not possible to tell from the difference scores whether changes are due to increases or decreases in the first or last halves of the interviews, the means for the first and last halves were plotted separately as a function of the successive sessions, and it was quite apparent that the between-group trends resulted almost entirely from changes in the last half of the interviews. In other words, the increasing tendency for the Regular group to approach further within interviews was not due to these *S*s starting off at a lower level of approach in the succeeding interviews, but from their obtaining increasingly higher approach ratings in the last half of the interviews.

The Feeling Inventory results indicate that the Regular and Nonverbal groups were inclined to report themselves as having experienced a larger decrease in anxiety after the five sessions than the Control group. An obvious possibility is that this self-report is not tapping "real" changes in anxiety level so much as *S*'s belief, that he should be less anxious, or that the experimenters expect him to be less anxious. On the other hand, the consistency of these results with the change in Meaning Unit rating scores and the GSR data suggest that they may be due to more than superficial changes in verbal report.

The only significant GSR finding was the between-groups quadratic component for the intra-interview change in skin conductance, which is illustrated in Fig. 2. This finding suggests that the Tape group showed an increasing tendency to become more anxious within interviews as the five sessions progressed, with this intra-interview increase in anxiety being especially marked in the last two sessions. The Regular group, on the other hand, showed increasing intra-interview changes in anxiety through the third session and thereafter showed progressively less of an increase in anxiety within interviews.

If this interpretation of the GSR data in terms of the hypothetical anxiety state of the individual is correct, it raises some interesting questions when one compares the GSR results with the Meaning Unit ratings of the verbal behavior. First, it should be noted that for both kinds of data it is the intra-interview changes that are significant, not the mean scores for the interviews. Second, there appears to be in part an inverse relationship between the GSR data and the Meaning Unit ratings. Thus, in the fourth and fifth sessions, the Regular group showed a decreasing amount of intra-interview change in anxiety at the same time that this group was showing the greatest amount of intra-interview increase in approach to emotionally important material. The Tape group, on the other hand, in these last two sessions showed a striking increase in intra-interview change in anxiety, at the same time that it was showing a decrease in the tendency to approach emotionally significant material. At first glance, these observations suggest that approaching emotionally important material is less anxiety arousing than avoiding it. But the situation is probably much more complex than that.

*S*s in the Regular group, as they began to verbalize emotionally important content, did indeed show increasing degrees of intra-interview increase in anxiety. These results suggest that the Tape *S*s were not completely withdrawing from the psychological field, but were experiencing concerns about which they were progressively more reluctant to talk. This suppressing behavior was apparently accompanied by a great deal of anxiety. If avoidance behavior on the part of the Tape group *S*s came to be associated with greater increases in anxiety, then one would expect a negative correlation between intra-interview changes in Meaning Unit ratings and skin conductance scores for these *S*s in the later interviews. The correlations between these two variables for interviews 4 and 5 were $-.47$ and $-.31$ respectively, which for an N of 9 are not significant but are in the expected direction.

This interpretation, as well as the empirical findings are consistent with evidence from a different study. Gordon, Martin, and Lundy (1959) induced conflict areas hypnotically, and then by posthypnotic suggestion caused the *S*s in one group in a subsequent interview first to "repress," then to suppress, and then to verbalize overtly the conflict material. In a second group, order was partly controlled by having them first "repress," then overtly verbalize, and then suppress. When skin conductance for the suppressing condition was compared with skin conductance for the verbalizing condition for the combined groups, it was found that *S*s showed a greater increase in conductance when suppressing than when overtly verbalizing. Both studies, then, support the hypothesis that approaching and verbalizing conflict areas is associated with less of an increase in anxiety than when overt discussion of troubling material is avoided.

More light may be shed on this relationship if a more refined analysis of the data were carried out. There is considerable evidence that the GSR is positively related to momentary changes in the therapy behavior of patients (Dittes, 1957; Panek & Martin, 1959), and it is doubtful if the present findings should be interpreted as indicating a negative relationship between momentary changes in approach behavior and GSR measures. It may be that there is some initial rise in conductance or reactivity as *S* first approaches a new therapy theme, followed then by an overshadowing decrease.

SUMMARY

From a population of college students who scored high on the Forced-Choice Manifest Anxiety Scale and who also volunteered for a brief psychotherapy experience, 27 *S*s were selected on the basis of an initial screening interview and assigned to three treatment groups. One group talked to a tape recorder with no therapist present; a second group experienced a therapist who responded by nonverbal means only, and a third group received regular psychotherapy.

*S*s were seen for five sessions with continuous tape and GSR recordings being obtained. A self-report index of anxiety, the Feeling Inventory, was also obtained before and after the five sessions. The verbal content of the interviews was rated in terms of degree of approaching or avoiding emotionally important areas.

The major significant findings were that the Regular group showed a tendency to increase in their intra-interview approach to emotionally important material as the five interviews progressed, whereas the Tape group showed the opposite trend, with the Nonverbal group lying in between. The GSR data indicated that the Regular group showed concomitant increases in anxiety as they approached emotionally important material in the first three sessions, but in the last two sessions their even further approach behavior was accompanied by less and less of an increase in anxiety. The Tape group, on the other hand, while showing a decreasing amount of intra-interview approach behavior as the sessions progressed, showed a marked increase in intra-interview change in anxiety as the sessions progressed. The Nonverbal group showed intermediate trends. The Feeling Inventory results indi-

cated that the Regular and Nonverbal groups reported a greater decrease in reported anxiety than a Control group, but the Tape group did not differ significantly from the Control group on this variable.

REFERENCES

DITTES, J. E. Extinction during psychotherapy of GSR's accompanying "embarrassing sex statements." *J. abnorm. soc. Psychol.*, 1957, **54**, 187–191.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–9.

GORDON, J. E., MARTIN, B., & LUNDY, R. M. GSR's during repression, suppression, and verbalization in psychotherapeutic interviews. *J. consult. Psychol.*, 1959, **23**, 243–251.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

GRANT, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, **59**, 149–151.

HEINEMAN, C. E. A forced choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447–454.

MARTIN, B. The validity of a self-report measure of anxiety as a function of the time interval covered by the instructions. *J. consult. Psychol.*, 1959, **23**, 468.

MILLER N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the Behavior Disorders*. Vol. I. New York: Ronald, 1944.

MURRAY, E. J., & BERKUN, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, **51**, 47–56.

PANEK, D. M., & MARTIN, B. The relationship between GSR and speech disturbances in psychotherapy. *J. abnorm. soc. Psychol.*, 1959, **58**, 402–405.

(Received October 17, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

EFFECT OF INTERPERSONAL RELATIONSHIPS UPON VERBAL CONDITIONING[1]

ALLAN SAPOLSKY
Adelphi College

RECENT findings on the conditioning of human subjects have shown that verbal reinforcement by an experimenter can modify *S*s' verbal responses in a predictable fashion and without *S*s' awareness. These findings have been interpreted by Krasner (1958) as having a direct bearing upon the process of psychotherapy.

However, currently available conditioning experiments have not controlled for the possibility that a relationship between the personal characteristics of *S* and *E* could affect the conditioning process. Such a possibility is inherent in the fact that, in a therapy setting, both positive and negative transference relationships are frequently found between patient and therapist and are believed to affect the outcome of the psychotherapeutic procedure.

The transference relationship can be viewed as one in which *S* overtly likes (or dislikes) *E*, or as one in which *S* is responding to personality characteristics which may (or may not) be possessed in common by *E* and *S*. The positive or negative quality of such interpersonal relationships could, in turn, be expected to have related effects upon the responses of *S*s in a verbal conditioning situation.

Work by Back (1951) indicated that instructions which orient *S*s to like or dislike each other can alter the judged attractiveness of an individual. In turn, groups composed of "attracted" individuals make more effort to influence each other and are more susceptible to influence. Generalizing this finding to the verbal conditioning process, it can be hypothesized that *E* exerts more influence (stronger conditioning) on an "attracted" group than on a "nonattracted" one. The first experiment of the present study was performed to test this hypothesis.

In the actual therapeutic situation, however, the "attractiveness" (or "unattractiveness") of *E* for *S* is not customarily established on the basis of overtly available information. Rather, it appears to develop within the therapy setting as the result of an interaction of the personality characteristics of *S* and *E*. The work of Murphy (1947) and Schutz (1958) suggests that people are attracted to individuals whose personality characteristics they perceive as empathic to their own (compatible characteristics). Likewise, they find unattractive those individuals whose personality characteristics are perceived as incompatible with their own. If "attractiveness," based on the perception of subtle cues stemming from related personality characteristics of *S* and *E*, is a factor in the verbal conditioning situations, then "compatible" *E*s should produce stronger conditioning than "incompatible" *E*s. This was the hypothesis which the second experiment in this study was designed to test.

The work of Festinger (1957) on compliant behavior suggested that the physical presence of *E* might serve to reinforce the effects of whatever personality characteristics were operating in the acquisition phase of conditioning, and that such affects might be modified in *E*'s absence. For this reason, both experiments were extended to include an extinction series with *E* absent—a condition which parallels that of the patient upon leaving the therapy setting.

EXPERIMENT I

Method

Subjects. *S*s were 30 female college freshmen representing a homogeneous group of education majors just graduated from high school and living as a group on campus. *S*s were randomly assigned to receive instructions designating *E* as either attractive or unattractive. Fifteen *S*s received High Attractiveness instructions, the other 15 Low Attractiveness instructions.

Materials. A total of 160, white, 3 × 5, unlined index cards were used as the stimuli. In the center of each card, a different commonly used verb was typed in the past tense. Six personal pronouns, "I," "We," "He," "She," "They," and "You" were typed below each verb, and their respective orders were randomized for the entire series of 160 cards. The stimulus cards were shuffled after each administration in order to randomize the order of presentation for each *S*.

---

[1] This paper is based on a dissertation submitted to the faculty of Adelphi College in partial fulfillment of the requirement for the degree of doctor of philosophy. The writer is deeply indebted to Marvin A. Iverson, under whose supervision the study was conducted; Dorothy Cantor, for serving as *E* in Experiment I; and Mary E. Reuder, for invaluable assistance in the preparation of this article.

Experimenter. E was a female graduate student whose physical appearance was such that she could be easily mistaken for an undergraduate student. She was dressed in a manner appropriate for undergraduate students on the campus. The same *E* was used for all *S*s.

Procedure. Just prior to her introduction to *E*, each *S* was informed by the investigator that this was an experiment which required the two persons who would be working together to be congenial. *S* was further told, "You will be paired with an experimenter whom you will find congenial. We have developed a questionnaire which enables us to do this."

The FIRO-B, a scale requiring ratings of behavior which one desires to find in social situations, was then administered and scored in *S*'s presence.

In addition, *S* was asked, "Can you describe the person you want to work with? What would be the most objectionable traits in a person with whom you would work?" *S* was then randomly assigned to either one of the experimental conditions.

Under the condition of High-Attraction, *S* was told, "Usually, we can't match people exactly, but in your case this will be possible." *S*s in the Low-Attraction group were told, "Usually, we can match people quite well, but in your case we're going to have some trouble. It's going to take too long to locate somebody for you, so I'm assigning you to Miss C. She may irritate you a little, but do the best you can."

After greeting and seating S, *E* read the following instructions which were adapted from those of Taffel (1955):

> You will see a word in the center of each card. I want you to make up a sentence using this word. Below the word in the center, you will see a group of other words. Take any one of these and use it to start your sentence (pause). Now, it doesn't matter whether the sentence you make up is long or short or even if it is complicated or simple. It is important that you answer with the first sentence that enters your mind. It isn't easy to do this, but you will find that, if you try to answer as quickly as possible, you are more likely to give the first thing that comes into your mind (pause). Everything we say is being recorded (point to microphone) on a tape recorder, so that it will not be necessary for me to write down your sentences now, but I will have a record of all the sentences you make up (pause). Do you understand the directions? All right, let's begin.

E turned each card after *S*'s response. In order to obtain a criterion level for the use of the critical pronouns "I" or "We", the responses of *S* to the first 20 cards were not reinforced. For the next three series of 20 cards, *E* provided reinforcement by saying "mmm-hmm" in a flat, unemotional tone at the end of any sentence which began with the "I" or "We" pronouns. When *S* had responded to the first 80 stimulus cards, the investigator, from behind a one-way vision screen, simulated a telephone ring in the experimental room. *E* answered the phone in the following manner, "Hello... yes... but I'm busy right now... all right." She then explained to *S* that she would have to leave, but that *S* should continue making up sentences until she returned. The extinction phase consisted of four series of 20 cards each, which *S* administered to herself while *E* was out of the room.

At the conclusion of the experiment, *S* was interviewed by the investigator. A series of questions, adapted from Tatz (1956), was asked to determine *S*'s degree of awareness of the conditioning procedure. Tatz had found a relationship between degree of awareness and rate of conditioning. Although, in the present study, all *S*s were aware of *E*'s use of "mmm-hmm," only three reported an association of this behavior with possible reinforcement effects. Two of these *S*s were from the Low-Attraction and one from the High-Attraction group.

As a check on the effectiveness of the instructions variable, *S*s were also asked to rate *E* on a self-anchored sociometric scale devised by Gardner and Thompson (1956). Of 15 *S*s receiving High-Attraction instructions, 11 rated *E* in the predicted direction, and 12 of those receiving Low-Attraction instructions responded in the manner predicted. A probability test indicated significance beyond the .01 level, verifying the effectiveness of the directions in establishing the two experimental groups.

Results

The curves of acquisition and extinction for both groups are plotted in Fig. 1. The *acquisition* data were then separately treated by analysis of variance according to Mixed Design I of Lindquist (1956). As can be seen in Table 1, the interaction between trials and attraction is significant beyond the .01 level. The nature of this interaction was further examined by applying a separate trend test to the slope of the curve of each attraction group.

FIG. I

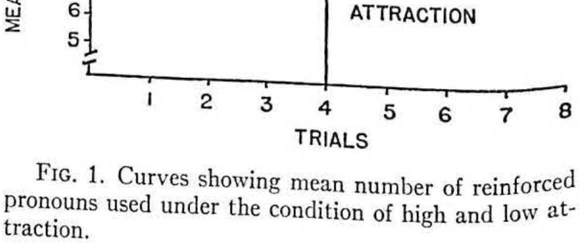


FIG. 1. Curves showing mean number of reinforced pronouns used under the condition of high and low attraction.

TABLE 1

Summary of Analysis of Variance for the Two Conditions of Attraction During the Acquisition Period

Source	*df*	Sum of Squares	Mean Square	*F* ratio
Total Between-*S*s	29	1125.5		
Attraction	1	154.1	154.1	4.43*
Error Between	28	971.4	34.7	
Total Within-*S*s	90	831.0		
Trials	3	195.3	65.1	11.03**
Interaction: T X A	3	140.8	46.9	7.93**
Error Within	84	494.9	5.9	
Total	119	1956.5		

* $p < .05$.
** $p < .01$.

TABLE 2

Summary of Analysis of Variance for the Two Conditions of Attraction during the Extinction Period

Source	*df*	Sum of Squares	Mean Square	*F* ratio
Total Between-*S*s	29	1847.9		
Attraction	1	29.0	29.0	<1.00
Error Between	28	1818.9	64.9	
Total Within-*S*s	120	1017.2		
Trials	4	26.5	6.6	<1.00
Interaction: T X A	4	218.0	54.5	7.90**
Error Within	112	772.7	6.9	
Total	149	2865.7		

* $p < .05$.
** $p < .01$.

The *F* ratio thus obtained for the High-Attraction group was significant beyond the .01 level. The Low-Attraction group did not reach the conventional level of significance.

The obtained interaction is therefore due to the fact that the High-Attraction group showed a continuously rising curve of acquisition of the CR, whereas the Low-Attraction group showed relatively little acquisition (see (Fig. 1). This finding confirmed the hypothesis that *S*s who find *E* attractive use the reinforced pronouns more frequently than *S*s who find *E* unattractive.

The extinction data were treated by analysis of variance, following the same design. Again, the interaction between trials and attraction was significant beyond the .01 level (see Table 2). However, examination of the extinction curves of the High-and Low-Attraction groups (Fig. 1) indicated that this interaction was not the same as that found for the acquisition curves. Trend tests of the slopes of the two extinction curves confirmed this difference. The *F* ratio for the slope of the High-Attraction group curve showed no significant departure from the horizontal base line, but the *F* ratio for the slope of the curve of Low-Attraction *S*s was significant beyond the .01 level. Thus, the Low-Attraction group may be said to have shown the effects of conditioning during the "extinction" period rather than during the "acquisition" phase of the experiment.

The "extinction" data of the High-Attraction group were further analyzed to determine whether experimental extinction actually occurred. The means of each of the extinction trials were separately compared with the mean of Trial 4 (the final acquisition trial). No significant decrease from the final acquisition level was found. The High-Attraction group retained the CR level of its original conditioning throughout the extinction period.

EXPERIMENT II

Method

Experimenters and subjects. Since the purpose of the second experiment was to study the effects of interpersonal relationships upon conditioning, it was necessary to hold constant the relevant personality characteristics of *E* and *S*. This was done by selecting *E* and *S* on the basis of their score patterns on the FIRO-B Scale, which measures needs in interpersonal relationships (Schutz, 1958). A preliminary study (Sapolsky, 1958) had shown this scale to be effective in establishing compatible and incompatible groupings among individuals having only classroom contact with each other.

Five female *E*s were selected from an undergraduate psychology class on the basis of their FIRO-B scores. These *E*s varied in their scores on the Inclusion and Affection areas, but all scored high on the Control area. *E*'s role in the conditioning procedure was essentially a controlling one; it was therefore deemed preferable to choose persons who desired to control interpersonal relationships.

*S*s for each *E* were then obtained by screening a variety of undergraduate classes in which *E* was enrolled. Approximately 300 female students were given the FIRO-B, from whom 30 were selected as suitable *S*s for this experiment. *S* was selected only if both she and her respective *E* were compatible (or incompatible) on all three of the interpersonal need areas measured by the FIRO-B (the method of establishing compatibility of need area has been derived by Schutz [1958]). Each *E* was assigned to 3 Compatible and 3 Incompatible *S*s with whom she worked individually. This procedure resulted in two groups of 15 *S*s each—one being highly compatible (Compatible Group) and one being highly incompatible (Incompatible Group) with its *E*s.

Since the *S*s were drawn from a relatively heterogeneous population, additional control data were gathered. No significant differences between the two groups of *S*s were found in age, intelligence, extent of acquaintance with *E*, or Taylor Manifest Anxiety Scale score.

Procedure. The procedure was identical to that in Experiment I, except that the "attractiveness" instructions were omitted. The awareness questionnaire and the self-anchored sociometric scale were again used, with similar results, as controls on the effectiveness of other procedures.

All but five *S*s (2 Compatible and 3 Incompatible) were aware of the "mmm-hmm" behavior of their *E*. Six *S*s (3 from each group) recognized the reinforcing effect of *E*'s responses. There was no systematic relationship between awareness of conditioning and the Compatibility-Incompatibility variable.

It had been predicted that Compatible *S*s and *E*s would like each other and that Incompatible *S*s and *E*s would not. On the sociometric scale, 22 of the 30 *S*s rated and were rated by their *E*'s in the predicted direction. Of the 30 pairs created by combining *E*s and *S*s, 19 rated each other as anticipated. A probability test showed these results to be significant beyond the .001 level. This finding established a similarity between the *relationships* studied in Experiment I and Experiment II despite the differences in operations employed.

Results

The data for Experiment II are graphically illustrated in Fig. 2. It can be readily seen that there is high similarity between these curves and those found in Experiment I (Fig. 1).

Separate analyses of variance of the acquisition and extinction data were carried out according to Lindquist's (1956) Mixed Design III.

FIG. 2

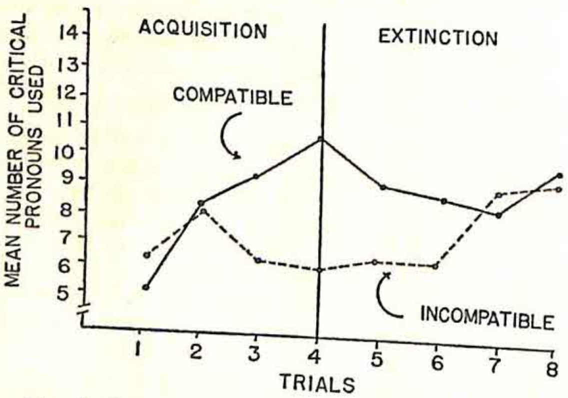


FIG. 2. Curves showing mean number of reinforced pronouns used by the compatible and incompatible groups.

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE FOR THE TWO COMPATIBILITY GROUPS DURING THE ACQUISITION PERIOD

Source	*df*	Sum of Squares	Mean Square	*F* ratio
Total Between-*S*s	29	574.9		
Compatibility	1	83.4	83.4	4.44*
Experimenters	4	46.2	11.6	<1.00
Interaction: C X E	4	69.3	17.3	<1.00
Error Between	20	376.0	18.8	
Total Within-*S*s	90	725.0		
Trials	3	152.7	50.9	11.84**
Interaction: T X C	3	161.4	53.8	12.51**
Interaction: T X E	12	78.7	6.6	1.53
Interaction: T X C X E	12	77.5	6.5	1.51
Error Within	60	254.7	4.3	
Total	119	1299.9		

* $p < .05$.
** $p < .01$.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE FOR THE TWO COMPATIBILITY GROUPS DURING THE EXTINCTION PERIOD

Source	*df*	Sum of Squares	Mean Square	*F* ratio
Total Between-*S*s	29	1423.0		
Compatibility	1	119.8	119.8	2.36
Experimenters	4	75.1	18.8	<1.00
Interaction: C X E	4	211.6	52.9	1.04
Error Between	20	1016.5	50.8	
Total Within-*S*s	120	1034.0		
Trials	4	81.9	20.5	2.85*
Interaction: T X C	4	132.2	33.1	4.59**
Interaction: T X E	16	44.7	2.8	<1.00
Interaction: T X C X E	16	197.7	12.4	1.72
Error Within	80	577.5	7.2	
Total	149	2457.0		

* $p < .05$.
** $p < .01$.

The acquisition data (Table 3) show the same kind of interaction (significant at the .01 level) as was found in Experiment I. Trend tests indicated that the Compatible Group had a significantly rising curve of acquisition, whereas the Incompatible Group (with one exception) did not. In the latter group, there was a significant rise on the first reinforced trial which was not sustained in later trials. These findings confirm the hypothesis that Compatible groups increase their use of reinforced pronouns more than Incompatible Groups. They further argue that similar effects on the acquisition of a conditioned verbal response occur when personal attractiveness is based on overt instructions as when attraction results from compatibility in interpersonal needs.

The extinction data in Experiment II also confirm the results of Experiment I. The interaction between trials and compatibility (Table 4) is significant beyond the .01 level. As in Experiment I, trend tests indicate that the increase in response of Incompatible *S*s accounts for the interaction. Compatible *S*s showed no appreciable over-all change over the acquisition phase. Thus, Incompatible *S*s in this experiment responded in the same manner as did Unattracted *S*s in Experiment I. They showed the effects of conditioning during the extinction period.

The extinction data of Compatible *S*s in Experiment II showed one major difference from that of Attracted *S*s in Experiment I. When the means of the extinction trials for Compatible *S*s were separately compared to the mean of their Trial 4 (maximum acquisition trial) a decrease, significant at the .05 level, was found for Trials 6 and 7. Nevertheless, the total amount of conditioning was still significantly above the initial criterion level.

DISCUSSION

The results of the two experiments support the major hypothesis that the positive or negative qualities of the interpersonal relationship between *S* and *E* have related effects upon *S*'s performance in a verbal conditioning situation. When there was either High-Attraction or Compatibility between *S* and *E*, the reinforcing value of "mm-hmm" was enhanced, presumably by the additional reinforcing qualities that *E* represented, and a typical learning curve was obtained during acquisition. When the Attraction between *S* and *E* was low (or *S* and *E* were Incompatible), no significant increase in the use of the reinforced pronouns occurred during the acquisition period of either experiment. The personal quality of *E* can here be viewed as having taken on the aspects of an aversive stimulus which resulted in a decrement in the observed strength of the conditioned response. Data from the extinction phase of the experiments also supported this interpretation. With the removal of the aversive stimulus (*E*), the effective quality of the "mmm-hmm" reinforcement was displayed in the form of a significant increase in the previously rewarded behavior. Thus, the effect of a Nonattractive or Incompatible *E* was to suppress or counteract the immediate effectiveness of the positive reinforcement. This would suggest that when results derived from conditioning studies are generalized to areas involving a relationship between two persons, the quality of that interpersonal relationship must be considered.

These results extend and qualify the finding of Back (1951), that in more highly attracted groups, the members are more receptive to influence. An important factor in Back's study was that *S*s could terminate face-to-face contact with each other at their own discretion, and thus did not feel forced to remain in the influencing situation. *S*s in the present study were required to remain in the influencing situation. The obtained results support Back's finding that less attracted *S*s are less willing to receive influence. However, it was also found that when *S*s were required to expose themselves to influence, both groups ultimately showed similar susceptibility to the effects of the influence, although these effects were delayed in the resistive group.

This finding suggests that Back's concept of "receptiveness to influence" needs to be refined to point up two independent possibilities—not only a willingness on the part of *S* to receive influence, but also *S*'s susceptibility to such influence once exposed. The results of the extinction phases indicate that individuals may actually become influenced but are unable or unwilling to display the effects of this influence within the social setting. Thus, *S*s who fail to give anticipated responses may actually be suppressing the effects of the major treatment until they leave the influential setting.

Two differences between the results of Experiments I and II should be noted. First, the initial significant rise in the curve for Incompatible *S*s did not occur in the curve for the Low-Attraction group; and second, the curve for the Low-Attraction group of *S*s rose sharply at the beginning of the extinction period, whereas a comparable rise for *S*s of the Incompatible group was delayed until the latter half of the extinction trials. These discrepancies probably resulted from the nature of the experimental procedures. The instructions to Low-Attraction *S*s immediately focused their perceptions on a dislike of *E*; the emotional situation was explicitly labeled for them. There was no such direct focusing of attention for Incompatible *S*s. Instead, their attitudes toward their *E*s were necessarily of an intrinsic nature and therefore could not be as quickly and as overtly evident to them. It had to develop as a result of social interaction between *E* and *S*.

Krasner (1958) has pointed out the similarity between the verbal reinforcement by *E* in conditioning procedures and the therapist's use of "mmm-hmm" in the psychotherapy setting. The findings of the present study suggest that the subtle cues provided by a therapist's use of "mmm-hmm" are likely to be effective only when the interpersonal relationship between him and the patient is positive or compatible. The results further suggest that, for resistive or negativistic patients, an accurate evaluation of psychotherapeutic effectiveness may not be possible at the time therapy is actually taking place. It may well be that for such patients, the influence experienced during treatment will be expressed only after termination of the therapeutic relationship.

SUMMARY

An investigation was designed to study the effects of interpersonal relationships upon verbal conditioning. Two conditioning experiments were conducted in which *E* used "mmm-hmm" to reinforce sentences which *S* began with pronouns "I" or "We". Experiment I relied upon instructions to create a condition of either High- or Low-Attraction between *S* and *E*. Experiment II utilized the FIRO-B scale to establish Compatible and Incompatible groups in terms of interpersonal needs. *E* was present during the acquisition phase and absent throughout the extinction period.

*S*s who were Incompatible, or not attracted to *E*, suppressed the effects of reinforcement until removed from *E*'s presence. The implications of these findings for generalizing from verbal conditioning situations to the patient-therapist relationship in clinical settings was discussed.

REFERENCES

BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

FESTINGER, L. *A theory of cognative dissonance.* New York: Row, Paterson, 1957.

GARDNER, E. F., & THOMPSON, G. G. *Social relations and morale in small groups.* New York: Appleton-Century-Crofts, 1956.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

LINDQUIST E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

MURPHY, G. *Personality.* New York: Harper, 1947.

SAPOLSKY, A. Conditioning of personal pronouns as a function of subject-experimenter compatibility and of the attractiveness of the experimenter to the subject. Unpublished doctoral dissertation, Adelphi College, 1958.

SCHUTZ, W. C. *FIRO-B: A three-variable theory of interpersonal relations.* New York: Rinehart, 1958.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

TATZ, S. J. Symbolic mediation in "learning without awareness." Unpublished doctoral dissertation, Yale Univer., 1956.

(Received October 3, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

OPERANT CONDITIONING IN A QUASI-THERAPY SETTING[1]

J. MAURICE ROGERS
Stanford University

RECENT studies of verbal conditioning, although not designed as investigations of the process of psychotherapy, have important implications for the understanding of interactions between the therapist and the patient. Conducted along the lines of Skinner's operant conditioning paradigm with rats and stimulated by Greenspoon's pioneering study (1955), subsequent experimenters have amply demonstrated that human verbal behavior can be operantly conditioned, that such conditioning can be accomplished without awareness, that a large variety of verbal behaviors may constitute Skinnerian "responses" and an equally large variety of environmental events serve as "reinforcing stimuli," and that typical conditioning phenomena (positive and negative reinforcement, discrimination, extinction, etc.) analogous to that found in animal experimentation can be produced.

Psychotherapists, whatever their theoretical affiliation, agree that psychotherapy implies effort directed toward the production of personality changes and is successful to the extent that specified changes occur. Although changes have been claimed in a variety of classes of behavior, such as anxiety, aggression, and dependency, the bulk of such claims have been unsupported by objective data. Among the more systematic studies are those of Rogers (1951) and his associates (Rogers & Dymond, 1954), who have repeatedly shown that an important change that occurs with psychotherapy is modification of the patient's self-concept. Raimy (1948), who was the first to demonstrate a decrease in negative and an increase in positive self-references in successful therapy, construed this as evidence of a reorganization of the self, and this view has since been strongly advanced by the Rogerian school. Reinforcement theory, however, makes tenable an alternative explanation—that the reported changes in self-reference verbalizations are elicited via unintentional selective reinforcement by the Rogerian therapist.

Research in operant conditioning of verbal behavior is gradually approaching the natural clinical setting of psychotherapy itself. In order for experimental findings to have maximal relevance for both the process and outcome of psychotherapy, it is essential that conditioning investigations be conducted in situations paralleling as closely as possible those of psychotherapy itself, yet retaining the controls and rigor essential to objective scientific study. Under such conditions, research findings will be more directly applicable to psychotherapy as it is characteristically conducted without the necessity of translating them into other terms and another frame of reference with the consequent hazard of overextrapolation. Manipulation of actual therapy itself, however, is limited by the clinician's ethical responsibility toward the patient, which compels him to exercise restraint in introducing conditions which conceivably might produce adverse effects in the patient. It was with these considerations in mind that the present experiment was designed and carried out. The hypotheses tested were, that an interviewer can produce changes in a subject's self-reference verbalizations by consistently reinforcing a particular class of such statements with simple stimuli, and that such reinforcements can alter a subject's concept of himself, as measured by personality tests.

METHOD

The experiment consisted of setting up a quasi-therapy situation involving a series of free-interviews with individual *S*s in which all self-reference verbalizations were reinforced by the interviewer with a simultaneous "Mm-hm" and nod of the head. Each *S* was told that the interviewer was making a study to learn how people think and feel about themselves and was asked to describe spontaneously his personality characteristics and traits without questions or comments from the interviewer. Every interview was tape recorded so that content analyses could subsequently be done directly from the tapes.

*S*s were 36 male students drawn from the introductory psychology course at Stanford University. They were randomly assigned to two experimental groups and a control group—Group A, reinforced for *positive* self-references; Group B, reinforced for *negative* self-references; and Group C, not reinforced at all. A total of 216 ten-minute "therapy" interviews (6 per *S*) were tape recorded, and a detailed content analysis of all self-references was made to ascertain if any relationship existed between self-references[2] and reinforcements.

---

[1] This report is based on the author's dissertation for the degree of doctor of philosophy, completed at Stanford University June 1958 under a committee consisting of Albert Bandura (Chairman), C. Leland Winder, and Paul R. Farnsworth, whose counsel is gratefully acknowledged.

A battery of four personality tests was administered prior and subsequent to the therapy to see if changes would occur, and whether such changes would be related to self-reference changes in therapy. Two of these tests, Adjective Self-Descriptions and Sentence Completions, were designed by the experimenter to induce the appropriate "set" for the ensuing therapy interviews by eliciting self-referent ideation. The other two tests were utilized to provide measures of personality change. These tests were the 50-item form of the Taylor Scale of Manifest Anxiety (Taylor, 1953) and the Q Sort Emotional Adjustment Test. The latter is a universe of 100 self-reference statements developed by Haigh and Butler and converted by Dymond into an adjustment test (Rogers & Dymond, 1954, p. 76).

Finally, a structured interview, likewise tape recorded, was conducted to ascertain *S*s' awareness of the conditioning and their reaction to the total experiment. All *S*s were individually tested and interviewed, four 1-hr. sessions being conducted with each *S*.

In scoring all the self-reference categories in the interviews and in the Sentence Completion Test, reliabilities were established by correlating ratings of two independent judges. Product moment reliability coefficients ranged from .83 to .95. In scoring self-references in the Adjective Self-Description Test, coding was done by three independent judges, the code assigned each item being based on majority agreement of the judges. Agreement was high, there being unanimity on almost three-fourths of the items.

RESULTS

Effect of Reinforcement on Self-References in Therapy

Table 1 shows the mean therapy interview scores for the three groups. The response classes signify positive, negative, and ambiguous self-references, other or non-self-references, and total number of responses. Interview 1 represents *S*'s *operant* level since the interviewer was silent throughout that interview. Interviews 2 through 6 were *conditioned* interviews for the experimental *S*s in Group A and Group B, but were *unconditioned* for the control *S*s in Group C.

TABLE 1

MEAN THERAPY INTERVIEW SCORES FOR THE THREE GROUPS

Response Class	P%	N%	A%	O%	Total
Experimental Group A					
Interview 1	27.31	33.77	27.07	11.77	47.92
Interviews 2–6	30.47	23.29	22.78	23.70	53.20
Change	3.16	−10.49	−4.29	11.93	5.28
Experimental Group B					
Interview 1	31.53	27.22	34.21	7.04	61.17
Interviews 2–6	22.01	37.47	27.19	13.33	61.15
Change	−9.52	10.25	−7.02	6.30	−.02
Control Group C					
Interview 1	33.17	33.07	25.37	8.40	54.00
Interviews 2–6	18.93	29.67	29.86	21.53	54.75
Change	−14.23	−3.39	4.49	13.14	.75

TABLE 2

t VALUES OF DIFFERENCES BETWEEN THERAPY INTERVIEW CHANGE SCORES

Response Class	P%	N%	A%	O%	Total Number of Responses
Group A vs. Group C	3.27^{a}	1.86	1.76	.30	1.18
Group B vs. Group C	.81	3.40**	2.50*	1.71	.23
Group A vs. Group B	3.45^{a}	5.24^{b}	.55	1.44	1.31

* Significant beyond the .05 level.

** Significant beyond the .01 level.

[a] Significant beyond the .005 level.

[b] Significant beyond the .0005 level.

TABLE 3

t VALUES OF DIFFERENCES BETWEEN OPERANT AND CONDITIONED RESPONSES

Response Class	P%	N%
Group A	1.02	1.60
Group B	2.90**	1.78*
Group C	2.99**	.59

* Significant beyond the .05 level.

** Significant beyond the .01 level.

In Table 2 the changes in each group are compared with those in the other groups and the significances of those comparisons determined by means of the *t* test.

In Table 3 the operant response level (Interview 1) is compared with the conditioned response level (Interviews 2 through 6) in each group and the significance of the change determined by the *t* test.

---

[2] "Self-reference" was operationally defined as a verbal response by *S* which describes him in some way, tells something about him, or refers to some affect he experiences. A self-reference may be positive (favorable), negative (unfavorable), or ambiguous (neither). The unit to be coded was termed a "response" and was defined basically as a grammatical sentence. This was a synthesis of "Sentence-Unit" and "Thought-Unit" scoring (Dollard & Mowrer, 1947), on which they achieved high rater reliability.

When the data are examined regarding the outcome of reinforcing *negative* self-references, a straightforward conditioning effect is found. From Table 3, it is seen that Group B, reinforced for negative self-references, showed a significant increase in that response class, while Groups A and C showed no significant change in that response. Table 2 shows that the change from operant to conditioned level in Group B was significantly different from the change in Groups A and C.

When the data are scrutinized for *positive* self-references, the picture is more complex. As shown in Table 3, Group A (reinforced for positive self-references) did not increase in that response class, and Groups B and C each showed a significant decrease in the response. From Table 2, however, it is seen that the change from operant to conditioned level in Group A is significantly different from the change in Groups B and C. It is clear, therefore, that the significance of the differences between the changes in the three groups is attributable to the decline in both Group B's and Group C's positive self-references rather than to the increase in Group A's positive self-references. Without reinforcement, positive self-references may tend to extinguish, whereas, with reinforcement, their extinction is arrested. The prediction that the interviewer's reinforcement of the *S*'s positive self-references would increase their frequency is thus only obliquely confirmed.

Relationship of Conditionability to Initial Level of Anxiety and Emotional Adjustment

It was predicted that conditionability (i.e., the readiness with which the *S* responded to reinforcement) is positively correlated with initial level of anxiety and negatively correlated with initial emotional adjustment. In other words, anxiety and poor emotional adjustment should facilitate the conditioning process. Rank difference correlation coefficients were computed between the Taylor Anxiety Score and conditionability, and between the Q Sort Adjustment Score and conditionability. The data did not support the hypothesis—no significant relationship was found between conditionability and anxiety, or between conditionability and emotional adjustment for either experimental group.

Effect of Therapy Conditioning on Personality Tests

Reduction in anxiety and improvement in emotional adjustment as a result of the conditioning in therapy was predicted. Accordingly, pre- and posttherapy scores were compared on the Taylor Scale and on the Q Sort Adjustment Test. No significant changes were found. It was noted, nonetheless, that on the Q Sort both experimental groups showed improved adjustment, the change in each group, however, not being significantly different from that in the control group. This suggests that with a longer period of conditioning—i.e., more or lengthier interviews—adjustment scores may be significantly increased.

To determine whether the conditioning of self-references *within* therapy would generalize to self-references *outside* of therapy, pre- and posttherapy scores were compared on the Adjective Self Description Test and on the Sentence Completion Test. Again, no significant changes were found.

Thus, it may be concluded from the data that the effect of the reinforcing stimuli was confined to the interview itself.

Consistency of Personality in Interview and Tests

By correlating the *S*'s responses to the four psychological tests with his *unconditioned* responses in quasi therapy, i.e., in Interview 1, it was possible to measure the consistency of personality. As would be anticipated, the data shows considerable, though by no means complete, personality consistency. The well adjusted person, it was found, has less manifest anxiety and refers to himself in more favorable terms in the free-interview, as well as in adjectives and in sentence completions, than the less well-adjusted person. Similarly, the person who refers to himself most unfavorably in the free-interview talks least. These, of course, are hardly revelations, being common clinical observations by psychotherapists. It may be somewhat comforting to clinicians, however, to have objective confirmation of some of their views.

Experimental Ss' Awareness of the Conditioning

"Awareness" was operationally defined as an explicit statement by *S* that the interviewer responded with "Mm-hm" or with a nod whenever *S* made a positive (or negative) self-reference. From the recordings of the structured interviews, it was determined that not a single *S* was aware of the head-nod conditioning; no one even mentioned it. Since *S*s in the main tended to avoid looking at the interviewer during the therapy interviews, this is not surprising. As for the "Mm-hm" conditioning, only 2 of the 24 experimental *S*s met the operational criterion for awareness, although 23 of them reported that they had noticed the "Mm-hm." Most of the latter interpreted the sound as a positive, rewarding response, signifying one or more of the following: a sign of encouragement or approval from the interviewer; an indication that the interviewer was listening, interested in, in agreement with or understanding of what *S* was talking about; or a sign that *S* was "on the right track" in talking about his personality characteristics rather than about extraneous matters. In short, while denying more categorical influence by the interviewer and failing to connect the "Mm-hm" with the specific verbal response class which it followed, they described the "Mm-hm" as a kind of generalized reinforcer which, by encouraging or reassuring them, helped them to continue talking or to remain on the over-all topic of their personality.

A few *S*s interpreted "Mm-hm" idiosyncratically: one thought it meant different things at different times—that the interviewer agreed with him, wanted him to speed up, or wanted him to change the topic. Another thought that it sometimes was meant as agreement and encouragement to continue with the topic, sometimes as a sign that he should drop the topic, and sometimes "to distract me." A third felt that it occurred "in disconnected places." A fourth held that it was intended "to get me off balance and upset my thinking . . . you were trying to see if I was easily disturbed."

Some *S*s insisted that they were entirely uninfluenced by the interviewer's "Mm-hm." One, affirming that the "Mm-hms" had had no effect on him, asserted, "I was just talking what I felt. I didn't care whether you agreed with me or not." Another declared, "I think if you had walked out of the room and given me the mike and stop-watch and asked me to talk for ten minutes about myself, I would have said exactly the same thing."

The single *S* who did not report that he had even noticed the "Mm-hms" declared that the interviewer was paying no attention at all to what he was saying but was focusing solely on his bodily manifestations of emotional tension.

It will be remembered that in the therapy interviews with the control group, the experimenter made no response whatsoever to *S*'s verbalizations except the generalized response of attention. The twelve subjects in this group unanimously asserted that the experimenter had in no way influenced their verbalizations.

In their reaction to the experimenter, however, they fell into two subgroups: those who reported that they pretty much ignored his presence, and those who reported that they were sensitive to his presence. The latter felt the experimenter was observing them closely either to detect signs of "nervousness" or "to analyze" what they were saying. The former tended to describe themselves as so absorbed by their self-verbalizations that they rapidly became oblivious to the presence of the experimenter.

DISCUSSION

The major finding of this experiment, that *S*'s verbalizations are influenced by an interviewer's reinforcements, is consistent with the evidence already accumulating in human operant conditioning investigations. This finding has considerable relevance for psychotherapy since changes in patient verbalizations are characteristically adduced as proof of personality change via psychotherapy. The various schools of psychotherapy generally attribute such changes primarily to monadic, intrapsychic events in the patient; ignoring, denying, or minimizing the immediate influence of the therapist. Even when they do acknowledge it, they typically fail to specify how the therapist is instrumental in determining the nature of the material elicited from the patient or the directions in which the patient changes. Since both positive and negative self-references were shown to be modifiable by interviewer reinforcement, one may ask if such changes signify, for example, changes in the Self, as the Rogerians assert, or if they are correlated with more general changes in the patient's motivational and attitudinal characteristics? Can one expect changes in behavior, verbal or otherwise, outside therapy to be correlated with verbal changes in therapy?

The present experiment suggests that the influence of minimal reinforcements, such as "Mm-hm," is restricted to the interview itself, for there were no correlative alterations of noninterview self-references or of personality test scores. Thus, as Winder (1957) put it, "One line of thought would consider psychotherapy as a complex and not always efficient conditioning process in which the patient learns to talk differently and little else." The present data may be construed as supporting such a generalization. On the other hand, psychotherapists could argue that personality modification results only from intensive and prolonged therapy and is not to be expected from a few short interviews.

The finding that conditioning took place without awareness (all *S*s were unaware of the head-nod conditioning and all but two were unaware of the "Mm-hm" conditioning) confirms what most verbal conditioning studies have previously found. It thus appears that changes in verbal behavior can be produced without *S* knowing that his behavior has been manipulated.

Since the data did not show a relationship between conditionability and initial level of either anxiety or emotional adjustment, no light is cast on the question of what personality variables in *S* facilitate or impede verbal conditioning. The results regarding anxiety conflict with those of Taffel (1955), whose data confirmed the thesis that anxiety connotes heightened responsiveness to stimuli. He found a positive relationship between manifest anxiety and conditionability. The fact that his *S*s were hospitalized psychiatric patients while the present ones were "normals" may account for this discrepancy. Presumably, his *S*s had higher anxiety, and the mean level of anxiety here may not have been of sufficient magnitude to exercise a discriminatory effect on conditionability.

As Krasner stated (1955, p. 21), "It sounds like the simplest and most axiomatic thing to say about psychotherapy, but it would seem reasonable to assume that the factor of the therapist listening, paying attention, showing some interest, is a basic and indispensable variable of the therapeutic situation." From the experimenter's experience with the control group, however, where he sat silently, making no response throughout the therapy, and where the majority of the *S*s reported themselves so absorbed in their own verbalizations that they became oblivious of his presence, he is tempted to challenge this axiom. Certainly, it would be worth investigating whether or not *S*s could verbalize as productively in the absence of an "interviewer" and whether the character of such verbalizations would differ systematically from those made with an interviewer present.

An equally heretical notion is that a form of psychotherapy might be devised which did not require the "therapist" to acquire clinical knowledge and training but merely to learn how to administer simple reinforcements (like "Mm-hm") to specified classes of verbal response! If subsequent investigations were to disclose a significant relationship between conditioning of particular classes of verbal response in the interview and desirable changes in personality test scores, it might well be that a simple verbal conditioning technique could be advantageously employed for circumscribed therapeutic purposes.

It may be argued that the quasi therapy was more quasi than therapy and that, therefore, the findings have little but analogic meaning for psychotherapy. Certainly, the situation was not structured to *S* as therapy, but rather as a laboratory experiment in personality. The *S*s themselves had not, ostensibly, sought therapy. On the other hand, the task for the *S* was almost identical with that in psychotherapy; to talk about his characteristics and traits and how he felt about them. Moreover, the behavior of the interviewer—listening attentively, nodding his head, saying "Mm-hm" at times, or saying nothing—was characteristic behavior for all therapists at least some of the time. Furthermore, the experiment did what other verbal conditioning experiments have not done, that is, dealt with areas of conflict (negative self-references), which is genuinely relevent to psychotherapy.

The demonstration that experiments in operant conditioning can be performed in therapy-like situations suggests the extension of such experimentation to actual psychotherapy. Tentative beginnings in this direction have been made with psychotics (Lindsley & Skinner, 1954; King, Merrell, Lovinger & Denny, 1957). It seems proven that at least some aspects of the complex phenomena represented by psychotherapy are amenable to behavioral analysis and experimental manipulation. Further investigations utilizing operant conditioning designs are indicated since such experimental designs appear to have heuristic value.

SUMMARY

The purpose of this experiment was to test the hypotheses that an interviewer can produce changes in a *S*'s self-reference verbalizations by applying simple reinforcements, and that such reinforcements can alter a *S*'s concept of himself.

The experiment consisted of setting up a quasi-therapy situation by conducting a series of free interviews with individual *S*s in which operant self-reference verbalizations were reinforced with a simultaneous "Mm-hm" and head-nod. All interviews were tape recorded. A battery of four personality tests, two designed to induce "set" and two chosen as measures of anxiety and emotional adjustment, was administered prior and subsequent to the therapy. Finally, a structured interview, likewise tape recorded, was conducted to ascertain the *S*'s awareness of the conditioning.

*S*s were 36 male university students in an introductory psychology course, randomly assigned to two experimental groups (one conditioned for positive and the other for negative self-references) and to a control group (not conditioned). A total of 216 ten-minute therapy interviews (6 per *S*) were recorded and a detailed content analysis of self-references made. Scoring reliabilities were established by correlating ratings of independent judges. The results were as follows:

1. Operant conditioning of verbal behavior in a quasi-therapy situation was demonstrated by establishing, beyond the .01 level, that negative self-references could be increased by interpolating simple reinforcing stimuli and that such conditioning could be accomplished without the *S*'s awareness. It was further shown, beyond the .01 level, that whereas positive self-references without reinforcement tended to extinguish themselves in quasi-therapy, with reinforcement their extinction could be arrested, this too without the *S*'s awareness.
2. Conditionability was not related to initial level of either anxiety or emotional adjustment.
3. Successful conditioning of self-references in therapy did not modify self-references outside of therapy or alter scores on tests of anxiety or emotional adjustment, indicating that the effect of the reinforcing stimuli was confined to therapy itself.
4. Consistency of personality was shown by *S*s' unconditioned responses in therapy and in psychological tests.

REFERENCES

DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3–32.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

KING, G. F., MERRELL, D. W., LOVINGER, E., & DENNY, M. R. Operant motor behavior in acute schizophrenics. *J. Pers.*, 1957, **25**, 317–326.

KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19–25.

LINDSLEY, O. R., & SKINNER, B. F. A method for the experimental analysis of the behavior of psychotic patients. *Amer. Psychologist*, 1954, **9**, 419–420. (Abstract)

RAIMY, V. C. Self-reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153–163.

ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.

ROGERS, C. R., & DYMOND, ROSALIND (Eds.) *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WINDER, C. L. Psychotherapy. *Annu. Rev. Psychol.*, 1957, **8**, 309–330.

(Received January 5, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

FEAR OF FAILURE AND UNREALISTIC VOCATIONAL ASPIRATION[1]

CHARLES H. MAHONE
Purdue University

THIS study is concerned with the application of a theory regarding fear of failure to the problem of vocational choice. Atkinson's usage (1958b) is followed in regard to fear of failure, which is defined as a motivational disposition to be anxiously concerned about avoiding failure. This disposition has been measured by two kinds of instruments: (*a*) the self-report measure of achievement-related anxiety, such as the Mandler-Sarason Test Anxiety Scale (Mandler and Sarason, 1952) or the Alpert Achievement Anxiety Scale (Alpert, 1957); and (*b*) the scoring of thematic apperception stories for achievement imagery (McClelland, Atkinson, Clark, & Lowell, 1953). In the latter case, fear of failure is inferred from low n Achievement[2] scores, for achievement anxiety is likely to be higher than positive achievement motivation for this group (Atkinson, 1958b, p. 332).

Two of the major factors that determine a person's suitability for a particular vocation are ability and interest. Therefore, if one is to be considered realistic in his choice of occupation, he should have ability commensurate with that required for substantial success in the occupation, and his interest pattern should be congruent with that of most people in his chosen field. The thesis of this paper is that persons who are fearful of failure tend to be unrealistic in their vocational choice with respect to both ability and interest.

With respect to the factor of interest, at least two studies (Atkinson, 1953; McClelland & Liberman, 1949) have shown that persons with relatively high fear of failure (as inferred from low n Achievement) tend to avoid consideration of achievement-related information. Fearful persons should therefore show a relative lack of relevant information concerning the kinds of interest satisfaction to be found in the various occupational areas. They should thus be likely to choose occupations only remotely related to the kinds of gratifications that they desire and expect to find in their vocations.

As for the ability factor, two considerations underlie the prediction. As in the case of interest, the fearful person may be expected to lack information concerning his own ability and that required for his choice of occupation. The second reason is based on theory regarding the level of aspiration of the fearful person. Atkinson (1958b) has offered a theoretical model for predicting level of aspiration from the relative strengths of fear of failure and need for achievement. He has suggested that for persons with higher positive than negative achievement motive, the aroused motivation to approach success is at a maximum when the achievement task is of intermediate difficulty. For persons with relatively high negative achievement motive (i.e., fear of failure), the aroused motivation to avoid failure is also at a maximum when the task is of intermediate difficulty. If both a person's tendency to approach success and his tendency to avoid failure are simultaneously aroused, as they are in competitive achievement situations, his behavior when confronted by a task of intermediate difficulty clearly depends upon which motive is stronger.

Applying this theory to level of aspiration, the fearful person (more strongly motivated to avoid failure than to achieve success) should tend either to overaspire or to underaspire (i.e., to avoid the intermediate range of the risk continuum). This prediction has been confirmed in several studies of risk-taking behavior in the experimental laboratory on simple competitive games (Atkinson, 1958c; Atkinson & Litwin, 1958; Clark, Ricciuti, & Teevan, 1956; Litwin, 1958; McClelland,

---

[1] This article has been adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy at the University of Michigan (Mahone, 1958). The author wishes to thank John W. Atkinson for his support and guidance in the research and in the preparation of this report. The author is also appreciative of the grant made for this study by the H. H. Rackham School of Graduate Studies at the University of Michigan.

[2] The abbreviation n Achievement stands for the *need for achievement* or the *achievement motive.* These two terms are used interchangeably as synonymous throughout the text.

1958). As yet, however, there is no theoretical basis for differentially predicting under- vs. overaspiration.

Let us regard the level of ability required by a person's vocational choice as an index of his level of aspiration. The fearful person may then be expected to aspire either to an occupation which he feels very confident about attaining (underaspiration) or to an occupation which he considers very difficult to attain (overaspiration). That is, he should avoid the choice of an occupation that represents an intermediate risk of failure for him—the most "realistic" choice in the sense that it falls within the "most probable" range of achievement (Lewin, Dembo, Festinger, & Sears. 1944). It is in this sense that the person fearful of failure is likely to be "unrealistic" with respect to ability in his vocational choice.

The following hypotheses are therefore offered:

1. A person who is fearful of failure tends to choose a vocation that is unrealistic in terms of the discrepancy between his own ability and the ability judged suitable for the occupation to which he aspires.

This is the most general hypothesis regarding ability, and the confirmation of this hypothesis should constitute evidence consistent with the parts of the theory concerning risk-taking and/or lack of information.

2. A person who is fearful of failure tends to choose a vocation that is unrealistic in terms of the discrepancy between his own *perceived* ability and the level of ability that he *perceives* as being suitable for the occupation to which he aspires.

This hypothesis rules out consideration of amount and accuracy of information, and the confirmation of this hypothesis should constitute evidence consistent with the risk-taking model.

3. A person who is fearful of failure tends to be inaccurate in his estimation of his own ability.

This hypothesis rules out consideration of risk, and the confirmation of this hypothesis should constitute evidence supporting the assumptions concerning lack of information as a determinant of unrealistic vocational choice.

4. A person who is fearful of failure tends to choose a vocation that reflects a pattern of interests discrepant from his own.

Since interest does not involve risk-taking behavior, the confirmation of this hypothesis should constitute evidence supporting the assumptions concerning lack of information as a determinant of unrealistic vocational choice.

METHOD

Subjects

The *S*s were 135 male college students drawn from introductory courses in psychology, French, and mathematics at the University of Michigan. They were volunteers from a larger group of approximately 330 male *S*s who had participated in another research project which involved the administration of the thematic apperceptive measure of n Achievement and the Alpert Achievement Anxiety Scale. In addition, the present *S*s were asked to answer the Strong Vocational Information Blank and a vocational information questionnaire.

Vocational Information Questionnaire

This questionnaire consisted of a series of questions designed to tap various factors associated with the *S*'s vocational choice. The most relevant items of information for the present study were obtained from questions that asked *S* to state: (*a*) his vocational objective; (*b*) an estimate of the ability required for each of a number of listed occupations; (*c*) an estimate of his own general ability relative to that of his fellow students; (*d*) what his vocational choice would be if ability were not a limiting consideration; and (*e*) which occupation in the above-mentioned list of occupations resembled most closely his own vocational choice. On the basis of pretest information on the vocational choices of a sample of 200 college students, the list of occupations was formulated so as to include, if not the actual stated vocational choice, at least a closely related occupation to that chosen by each *S*. It was the *S*'s estimate of the level of ability required for this listed occupation which constituted the measure of estimated ability required for own vocational choice.

n Achievement Scale

Fear of failure has been defined conceptually as a relatively stronger disposition to avoid failure than to approach success. In order to assess this variable, it was necessary to use measures of both positive and negative achievement motivation. The n Achievement test, the measure of positive achievement motivation, consisted of a set of six pictures recently included in a national survey by the Survey Research Center at the University of Michigan. The pictures were selected so that the stories could be scored for the achievement, affiliation, and power motives (Atkinson, 1958a). The test was administered under neutral classroom conditions (McClelland et al., 1953). The stories were scored for n Achievement by three trained scorers, each of whom scored stories to two pictures. A check on the reliability of scoring was made by having another experienced scorer score all six stories for a group of 30 *S*s. The rank order correlation between the latter scorer's total scores and the combined scores of the three original scorers was +.80. The mean n Achievement score for the present sample was 5.0, and the best median cutting point for both the larger group as a whole and the present sample was between 4 and 5, which placed 67 *S*s above the median (High n Achievement) and 68 *S*s below the median.

Debilitating Anxiety Scale

The measure of achievement anxiety (motivation to avoid failure) used was Alpert's Debilitating Anxiety Scale (Alpert, 1957). This scale is a part of Alpert's more general Achievement Anxiety Scale, which was designed in an attempt to refine the Mandler-Sarason Test Anxiety Scale by differentiating debilitating and facilitating anxiety. An early validation study on this instrument showed the Debilitating Anxiety Scale to correlate +.79 ($N = 40$) with the Test Anxiety Scale ($N = 40$), and, like the latter, to correlate negatively with college grades and performance on achievement and aptitude tests (Alpert, 1957).

The Debilitating Anxiety Scale measures *S*'s awareness of the extent to which anxiety interferes with the efficiency of his performance in achievement test situations. An illustrative item is: "In a course where I have been doing poorly, my fear of a bad grade cuts down my efficiency." The scale consists of ten such items, and *S* is asked to respond by indicating, on a five-point scale, how often he experiences the described situation. A score of 50 represents the highest degree of debilitating anxiety. For the present sample scores ranged from 12 to 39 with a mean of 25.9, and the best median cutting point was between 25 and 26. The best median cutting point for the larger group, however, was between 26 and 27. Since the larger group was more heterogeneous and representative of the college population as a whole, the latter cutting point was used, yielding 59 *S*s in the present sample who scored above the median (High Anxiety) and 76 *S*s who scored below the median.

The product-moment correlation between n Achievement and Debilitating Anxiety was +.09 ($N = 135$). When *S*s were grouped according to their position above or below the median score on both n Achievement and Debilitating Anxiety, 31 *S*s were classified as High n Achievement-High Anxiety, 36 as High n Achievement-Low Anxiety, 28 as Low n Achievement-High Anxiety, and 40 as Low n Achievement-Low Anxiety.

RESULTS

Judgments of Unrealistic Vocational Choice (With Respect to Ability)

The first hypothesis requires some evaluation of the discrepancy between *S*'s general ability and the ability required in order to achieve his vocational objective. Because no adequate information was found in the literature concerning the level of ability required in various occupations, clinical judgments were employed as the major criteria in determining realism of vocational choice with respect to ability. Two clinical psychologists, both having considerable experience in vocational counseling, made independent judgments of the realism of the vocational choice of each of the *S*s on the basis of the following information: ACE[3] score (total of the verbal and quantitative scores expressed as a percentile), grade point average for the preceding semester, class, major field, vocational choice, and "ideal" choice (i.e., the vocational choice that would be made if ability were not a consideration). The information on "ideal" choice was included to permit judgments of "underaspiration" in cases where the student was judged to have sufficient ability for a particular vocation, even though he rejected it because of perceived lack of ability.

The judges were asked to assign each *S* to one of five categories: Probably Over-Aspiring, Perhaps Over-Aspiring, Probably Realistic, Perhaps Under-Aspiring, Probably Under-Aspiring. They were also asked to indicate whether they were Very Confident, Moderately Confident, Moderately Doubtful, or Very Doubtful about their assignments in each case. The investigator, who has had three years experience as a trainee in vocational counseling, made independent judgments for each *S* to provide a basis for resolving disagreements between the two expert judges.

The three judges were in agreement with one another in 71–81% of the cases in judging the *S* to be either "realistic," "overaspiring," or "underaspiring". With the investigator making the deciding judgment in case of disagreement between the two expert judges, the final ratings consisted of 80 *S*s judged to be realistic, 45 judged to be overaspiring, and 10 judged to be underaspiring. Each of the three judges agreed with these final ratings in at least 81% of the cases. In the analysis, ratings of overaspiration and underaspiration were combined in a single "unrealistic" category.

According to the first hypothesis, *S*s who are high in achievement motivation and low in anxiety about failure should be judged more realistic in aspiring to vocations that are commensurate with their ability than *S*s who are low in achievement motivation and high in anxiety. The relative frequency of realistic

[3] American Council on Education test of verbal and quantitative aptitude.

and unrealistic persons in the various motivational groups are all in the predicted direction (Table 1). A highly significant difference is found between the High n Achievement-Low Debilitating Anxiety group and the Low n Achievement-High Debilitating Anxiety group ($\chi^2 = 7.96$, $df = 1$, $p < .003$). Debilitating Anxiety alone also discriminates between Realistic and Unrealistic groups ($\chi^2 = 10.1$, $df = 1$, $p < .001$), but n Achievement alone does not.

TABLE 1

n ACHIEVEMENT AND DEBILITATING ANXIETY RELATED TO ALL CLINICAL JUDGMENTS OF REALISTIC AND UNREALISTIC VOCATIONAL CHOICE

($N = 135$)

n Ach	Anxiety	N	Clinical Judgments		Anxiety	N	Clinical Judgments	
			Realistic %	Unrealistic %			Realistic %	Unrealistic %
High	High	31	48	52	High	59	44	56
High	Low	36	75*	25*	Low	76	71	29
Low	High	28	39*	61*		$\chi^2 = 10.1$		
Low	Low	40	68	32		$p < .001$		
* $\chi^2 = 7.96$					n Ach			
$p < .003$					High	67	63	37
					Low	68	56	44
						$\chi^2 = 1.09$		
						$p < .15$		

* Significance levels reported in this table are based on one-tailed tests since the direction of the differences was predicted.

TABLE 2

n ACHIEVEMENT AND DEBILITATING ANXIETY RELATED TO CONFIDENT, AGREED-UPON CLINICAL JUDGMENTS OF REALISTIC-UNREALISTIC VOCATIONAL CHOICE ($N = 58$)

n Ach	Debil. Anxiety	N	Clinical Judgments		Debil. Anxiety	N	Clinical Judements	
			Realistic %	Unrealistic %			Realistic %	Unrealistic %
High	High	9	56	44	High	21	33	67
High	Low	18	94*	06*	Low	37	80	20
Low	High	12	17*	83*		$\chi^2 = 15.0$		
Low	Low	19	74	26		$p < .0005$		
* $\chi^2 = 15.5^a$					n Ach			
($p < .0005$)					High	27	81	19
					Low	31	52	48
						$\chi^2 = 5.68$		
						$p < .01$		

* Significance levels reported in this table are based on one-tailed tests since the direction of the differences was predicted.

[a] Yates correction was used in obtaining this chi square value because of the small N.

Also, since the clinical judgments are obviously a rather gross and imperfect measure, a more refined analysis was made using only ratings on which both expert judges were not only in agreement but rated themselves as being either "Very Confident" or "Moderately Confident." There were 58 judgments, including 38 "realistic", 19 "overaspiring", and one "underaspiring".

In this more rigorous analysis (Table 2), n Achievement by itself is significantly associated with judged reality of vocational choice ($\chi^2 = 5.68$, $df = 1$, $p < .01$). The degree of association between unrealistic aspiration and Debilitating Anxiety is also very much enhanced ($\chi^2 = 15.0$, $df = 1$, $p < .0005$), and the difference between the High n Achievement-Low Anxiety and the Low n Achievement-High Anxiety groups is striking ($\chi^2 = 15.5$,[4] $df = 1$, $p < .0005$).

It may be noted in Table 2 that when *S*s are classified according to both n Achievement and anxiety, the distinctive group is the one low in n Achievement and high in Anxiety, which is clearly more unrealistic than the other three groups combined ($\chi^2 = 12.0$[4], $df = 1$, $p < .0005$).

Subjective Goal Discrepancy

The second prediction was that *S*s who are fearful of failure should avoid choosing occupations that they perceive as requiring the same level of ability as they perceive in themselves. Instead they should favor occupations that they perceive as requiring either more or less than their own perceived ability. In order to test this hypothesis, it was necessary to obtain a measure of each *S*'s estimate of his own ability and his estimate of the level of general ability necessary in order to attain his occupational goal. On the initial questionnaire all *S*s were asked: "What percentage of the undergraduates at the University of Michigan do you feel have sufficient *general ability* (i.e., verbal and quantitative ability) to attain the following occupational goals, provided they were motivated to do so?" As a measure of their perception of their own ability, they were also asked: "Irrespective of your own grade point average, where do you think you stand in relation to your fellow students in *general ability*?" Both of these estimates were made relative to "fellow students" so that they would be comparable.

---

[4] Computed with Yates correction because of the small N.

The subjective goal discrepancy, then, is the difference between the *S*'s estimate of his own ability and his estimate of the ability required for his vocational objective. The formula for the discrepancy score is the perceived percentage of students with higher ability than *S* minus the perceived percentage of students with sufficient ability to attain *S*'s vocational goal. The range of discrepancy scores was from −48 (own ability seen as superior to that required by job) to +85 (own ability seen as inferior to that required by job). The median discrepancy score was +7 if the direction of discrepancy is taken into account. If only the absolute size of the discrepancy is considered, the median discrepancy is ±12.

This index is somewhat problematic, however, in that it raises the question of whether the best test of the hypothesis is made using the median cutting point which takes into consideration the direction of the discrepancy, or the median cutting point which disregards direction. If one assumes that the *S*s are responding to the question of estimating ability required by various occupations by estimating the *average* ability required, then the theoretical model for risk-taking behavior would certainly call for the cutting point which disregards the direction of discrepancy, since this estimate would be analogous to an estimate of intermediate risk, and thus large negative discrepancies would represent underaspiration. If however, the *S*s are making estimates in terms of the *minimum* ability required, then this kind of estimate provides no basis for determining underaspiration, for occupations probably differ markedly in the range of ability levels which they can productively use. In this case negative discrepancies would have to be regarded as realistic, regardless of size. Since the basis on which *S*s arrived at their estimates is not known, analysis of the data was made both considering and disregarding the direction of the discrepancy.

When the direction of the discrepancy was considered, the median cutting point was +8, in terms of which 71 *S*s were classified as "realistic", and 63 were classified as "unrealistic", which, in this case, means only overaspiring (i.e., positive discrepancies[5]). When the direction of the discrepancy was disregarded, the median cutting point was ±13, in terms of which 68 *S*s were classified as "realistic", and 66 were classified as "unrealistic", which, in this case, includes both overaspiring (positive discrepancies) and underaspiring (negative discrepancies) *S*s.

TABLE 3

n ACHIEVEMENT AND DEBILITATING ANXIETY RELATED TO POSITIVE SUBJECTIVE GOAL DISCREPANCY (OVERASPIRATION)

n Ach	Anxiety	*N*	Positive Goal Discrepancy High %	Positive Goal Discrepancy Low %
High	High	30	53	47
High	Low	36	31*	69*
Low	High	28	61*	39*
Low	Low	40	48	52
			*$\chi^2 = 5.69$ $p < .01$	

Anxiety	*N*	Positive Goal Discrepancy High %	Positive Goal Discrepancy Low %
High	58	57	43
Low	76	39	61
	$\chi^2 = 3.97$ $p < .03$		
n Ach			
High	66	41	59
Low	68	53	47
	$\chi^2 = 1.92$ $p < .10$		

* Significance levels reported in this table are based on one-tailed tests since the direction of the differences was predicted.

In the positive discrepancy analysis (Table 3), the hypothesis is confirmed for the combination of n Achievement and Debilitating Anxiety ($\chi^2 = 5.69$, $df = 1$, $p < .01$). Debilitating Anxiety alone is also significantly related to this index of overaspiration ($\chi^2 = 3.97$, $df = 1$, $p < .03$), but n Achievement alone is not. In the absolute discrepancy analysis (both over- and underaspirants considered unrealistic), none of these associations reached the .05 level of significance when the distribution of absolute discrepancies is divided at the median ±13. However, if we accept a more stringent definition of the "realistic" category which includes *only the least discrepant third of the distribution* of absolute goal discrepancies (±8), as suggested in a study by Atkinson and Litwin (1958), then Debilitating Anxiety alone ($\chi^2 = 5.36$, $df = 1$, $p < .02$) and the High n Achievement-Low Anxiety vs. Low n Achievement-High Anxiety comparison ($\chi^2 = 7.0$, $df = 1$, $p <$

---

[5] The total *N* here is only 134 because one *S* failed to provide an estimate of his own ability.

TABLE 4

n ACHIEVEMENT AND DEBILITATING ANXIETY RELATED TO ABSOLUTE SUBJECTIVE GOAL DISCREPANCY (INTERMEDIATE THIRD VS. COMBINED POSITIVE & NEGATIVE EXTREMES)

n Ach	Anxiety	N	Absolute Goal Discrepancy		Anxiety	N	Absolute Goal Discrepancy	
			Mid-third %	Extremes %			Mid-third %	Extremes %
High	High	30	30	70	High	58	24	76
High	Low	36	50*	50*	Low	76	43	57
Low	High	28	18*	82*			$\chi^2 = 5.36$	
Low	Low	40	38	62			$p < .02$	
					n Ach			
		$^*\chi^2 = 7.0$			High	66	41	59
		$p < .005$			Low	68	29	71
							$\chi^2 = 2.07$	
							$p < .10$	

* Significance levels reported in this table are based on one-tailed tests since the direction of the differences was predicted.

.005) show highly significant relationships with subjective goal discrepancy (Table 4). n Achievement alone does not reliably discriminate between the mid-third realistic group and the over- and underaspirant groups combined. However, the trends are all in the predicted direction, and it is very clear that the only group heavily represented in the "realistic" category is the High n Achievement-Low Anxiety group.

Inaccuracy in Estimation of Own Ability

In the initial questionnaire, *S*s were asked to estimate their standing in general ability in comparison with fellow students, in terms of percentile rank, which would then be directly comparable to *S*s' actual percentile scores for combined verbal and quantitative ability on the ACE. One measure of unrealism (or inaccuracy) is the discrepancy between the percentile ranks of the *S*'s estimate of his ability and his actual ability; the latter percentile rank was subtracted from the former. The ranks for measured ability ranged from the 4th to the 99th percentile, with a median at the 61st percentile. The percentile discrepancies ranged from −47 (underestimation) to +75, with a median absolute discrepancy of ±16. There were 35 negative discrepancies and 99 positive discrepancies[6]. The median cutting point was between ±15 and ±16, yielding 68 *S*s classified as "inaccurate" and 66 *S*s classified as "accurate" in estimating their own ability.

The third hypothesis states that *S*s with high achievement motivation and low anxiety are more accurate in estimating their own general level of ability than are *S*s with low achievement motivation and high anxiety. This prediction was confirmed ($\chi^2 = 3.52$, $df = 1$, $p < .04$). Debilitating Anxiety and n Achievement taken singly were not significantly related to accuracy of estimate (Table 5). The major difference in Table 5 is between the Low n Achievement-High Anxiety group and the other three groups combined ($\chi^2 = 4.6$, $df = 1$, $p < .02$).

Unrealistic Vocational Choice (With Respect to Interest)

At the same time that the judges made their rating of the realism of vocational choice with respect to ability, they also indicated the primary and secondary vocational areas on the Strong interest test which they felt would reflect the interest pattern most congruent with the *S*'s vocational choice. On the basis of the *S*'s scores in these designated areas, as evaluated in terms of the pattern analysis suggested and used by Darley (1941), plus his score on the specific occupational key related to his vocational choice, the investigator divided the *S*s into two approximately equal groups. One group ($N = 66$) was considered to be nondiscrepant in their interest patterns

TABLE 5

n ACHIEVEMENT AND DEBILITATING ANXIETY RELATED TO INACCURACY IN ESTIMATE OF OWN ABILITY

n Ach	Anxiety	N	Inaccuracy		Anxiety	N	Inaccuracy	
			High %	Low %			High %	Low %
High	High	30	46	54	High	58	57	43
High	Low	36	44*	56*	Low	76	46	54
Low	High	28	67*	33*			$\chi^2 = 1.41$	
Low	Low	40	47	53			$p < .15$	
					n Ach			
		$^*\chi^2 = 3.52$			High	66	45	55
		$p < .04$			Low	68	56	44
							$\chi^2 = 1.46$	
							$p < .15$	

* Significance levels reported in this table are based on one-tailed tests since the direction of the differences was predicted.

---

[6] The total *N* here is only 134 because one *S* failed to provide an estimate of his own ability.

TABLE 6

n ACHIEVEMENT AND DEBILITATING ANXIETY RELATED TO REALISTIC-UNREALISTIC VOCATIONAL CHOICE (WITH RESPECT TO INTEREST)

n Ach	Anxiety	N	Interest: Realistic %	Interest: Unrealistic %
High	High	31	61	39
High	Low	36	56*	44*
Low	High	28	32*	68*
Low	Low	40	45	55
		*$\chi^2 = 3.50$ $p < .04$		

	N	Interest: Realistic %	Interest: Unrealistic %
Anxiety			
High	59	47	53
Low	76	50	50
	$\chi^2 = .12$ $p < .40$		
n Ach			
High	67	58	42
Low	68	40	60
	$\chi^2 = 4.55$ $p < .02$		

* Significance levels reported in this table are based on one-tailed tests since the direction of the differences was predicted.

(i.e., inventoried interests were congruent with vocational choices) and the other group ($N = 69$) was considered to be discrepant (i.e., inventoried interests were not congruent with vocational choices).

n Achievement alone was found to be significantly related to congruence (or reality) of interest pattern in the expected direction ($\chi^2 = 4.55$, $df = 1$, $p < .02$), and a significant difference was also found between the High n Achievement-Low Anxiety group and the Low n Achievement-High Anxiety group ($\chi^2 = 3.50$, $df = 1$, $p < .04$). Anxiety alone, however, was not significantly related to realism of interest pattern (Table 6).

Relationships Among Indices of Unrealism

The relationships among the indices of unrealism about ability are all in the expected direction, and three of the five comparisons are statistically significant beyond the .05 level. The dimensions that do not appear to be significantly related are the Clinical Judgments and both the Positive and the Absolute Subjective Goal Discrepancy. The relationships between the unrealistic interest index and the ability indices are also all in the expected direction, but only one of the relationships with Interest, that of Absolute Goal Discrepancy, is statistically significant ($\chi^2 = 3.00$, $df = 1$, $p < .05$).

Ability, Fear of Failure, and Indices of Unrealism

The results show that measured ability (ACE percentile score) is significantly related (negatively) to Debilitating Anxiety ($\chi^2 = 11.8$, $df = 1$, $p < .001$), but is not related to n Achievement. Ability is also significantly associated with all of the indices of unrealism related to ability ($p < .005$ in all cases). In every case the group low in ability is typically more unrealistic than the group high in ability. However, ACE percentile score is not significantly related to Interest Reality.

DISCUSSION

Because of the association between high measured ability and low Debilitating Anxiety, it is very difficult to distinguish between the relationships of ability and those of motivation to the three ability-related dimensions of unrealistic vocational aspiration. This kind of distinction is desirable, however, in view of the possibility that the relationships of ACE score to Clinical Judgments and to Inaccuracy in Estimating Own Ability are artifacts resulting from the fact that measured ability was directly involved in arriving at these two indices.

In view of this possibility, it is important to determine whether or not fear of failure is related to Clinical Judgments and Inaccuracy in Estimating Own Ability independent of level of ability. For this reason, an analysis of covariance was accomplished to determine whether the average level of anxiety in "Realistic" and "Unrealistic" groups was different (as predicted) when the groups were equated by this statistical procedure on ACE score. The results clearly show that fear of failure is related to Clinical Judgments ($N = 135$) when ability is held constant ($F = 5.40$; $df = 1{,}132$; $p < .02$[7]). This means that it would *not* be possible to account for the relationship between fear of failure and judgments of unrealistic vocational aspiration solely in terms of a possibly artifactual relationship between judgments of unrealistic aspiration and ability. An analysis of covariance was not made using the Inaccuracy index to categorize realistic and unrealistic groups because an inspection of the chi square tables indicated that no relationship between this index and the motivation measures remained when *S*s high and low on the ACE were considered separately. Most of the relationship between this index and the fear of failure measure reported in Table 6 should be attributed to the relationship of both kinds of variables to the general ability measure.

---

[7] One-tailed test.

Let us now consider the relationship between low intelligence and high anxiety about failure, since the interaction of these two variables is probably of substantial importance in the prediction of unrealistic vocational aspiration. One reason why we might expect these variables to be related is that relatively low intelligence may cause a person to have more early failure experiences and thus to become more fearful of failure. A second and perhaps even more important factor is that anxiety about achievement has on numerous occasions been shown to interfere with performance on aptitude tests (Alpert, 1957; Sarason & Mandler, 1952; Zweidelson, 1956). Since our measure of ability, the ACE test, is just such a performance test taken under competitive conditions, we should expect that high anxiety would lead to performance decrements, i.e., lower ACE test scores. For this reason, when we control for measured ability in the analysis of covariance we are probably overcontrolling. Not only are we ruling out the effect of innate ability on fear of failure, but we are in all likelihood also ruling out the effect of fear of failure on the performance measure of innate ability. Thus, we have good reason to believe that the relationship of fear of failure to unrealistic vocational aspiration, independent of the effect of innate ability, is even stronger than the results of the analysis of covariance show it to be.

The results of the present study confirm the applicability of the principles found in the laboratory study of level of aspiration to aspiration in real life.

Summary

The purpose of this study was to determine whether fear of failure in college students is a determinant of unrealistic vocational aspiration.

Vocational aspiration is considered in the context of theory having to do with motivational determinants of level of aspiration. Persons who are fearful of failure are presumed to be generally avoidant in their behavior in competitive achievement situations. On the one hand, such a person should avoid consideration of achievement related information, and, on the other hand, he should prefer speculative ventures, where his probability of success is quite low (overaspiration), or safe ventures, where his probability of success is quite high (underaspiration). In contrast, persons who are relatively strong in motivation to achieve success should tend to prefer ventures where the probability of success is intermediate.

Measures of both achievement motive and achievement-related anxiety were used in the present study. The *S*s were 135 male college students. They were classified as "realistic" or "unrealistic" in vocational aspiration in terms of each of four criteria: (*a*) judgments by clinical psychologists based on the discrepancy between the *S*'s own measured ability and the ability judged to be required by the *S*'s vocational choice; (*b*) the discrepancy between the *S*'s estimate of his own ability and his estimate of the ability required to attain his vocational goal; (*c*) the *S*'s inaccuracy in estimating his own ability; and (*d*) the discrepancy between the *S*'s interest pattern and his vocational aspiration. On each criterion of realistic vs. unrealistic vocational aspiration, significantly more *S*s who were low in achievement motivation and high in achievement-related anxiety were classified as unrealistic, than *S*s who were high in achievement motivation and low in achievement-related anxiety. Thus, the major hypothesis was clearly supported.

REFERENCES

Alpert, R. Anxiety in academic achievement situations: Its measurement and relation to aptitude. Unpublished doctoral dissertation, Stanford Univer., 1957.

Atkinson, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

Atkinson, J. W. (Ed.) *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. (a)

Atkinson, J. W. Motivational determinants of risk-taking behavior. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society.* Princeton: Van Nostrand, 1958. Pp. 322–339. (b)

Atkinson, J. W. Toward experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 288–305. (c)

Atkinson, J. W., & Litwin, G. H. Achievement motive and test anxiety as motives to approach success and to avoid failure. Unpublished manuscript, Univer. of Michigan, 1958.

Clark, R. A., Ricciuti, H. N., & Teevan, R. C. An investigation of the hope of success and fear of failure aspects of need for achievement. *J. abnorm. soc. Psychol.*, 1956, **53**, 182–186.

Darley, J. G. *Clinical aspects and interpretation of the Strong Vocational Interest Blank*. New York: The Psychological Corporation, 1941.

Lewin, K., Dembo, Tamara, Festinger, L., & Sears, Pauline S. Level of aspiration. In J. McV. Hunt (Ed.), Vol. 1, *Personality and the behavior disorders*. New York: Ronald Press, 1944. Pp. 333–378.

Litwin, G. H. Motives and expectancy as determinants of preference for degrees of risk. Unpublished honors thesis, Univer. of Michigan, 1958.

Mahone, C. H. Fear of failure and unrealistic vocational aspiration. Unpublished doctoral dissertation, Univer. of Michigan, 1958.

Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

McClelland, D. C. Risk taking in children with high and low need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 306–321.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century, 1953.

McClelland, D. C., & Liberman, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.

Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.

Zweidelson, I. Test anxiety and intelligence test performance. *J. consult. Psychol.*, 1956, **20**, 479–481.

(Received March 2, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

THREE VERSIONS OF THE F SCALE AND PERFORMANCE ON THE SEMANTIC DIFFERENTIAL[1]

ROBERT E. MOGAR
State University of Iowa

IN RECENT years inferences of authoritarian personality trends from the California F Scale have been challenged by a number of investigators. Considerable evidence has accumulated to suggest the importance of determinants of F scale performance other than authoritarian attitudes. Particular attention has been called to acquiescence response set because of unidirectional wording and scoring of the scale (Bass, 1955; Chapman & Campbell, 1957; Christie, Havel, & Seidenberg, 1958; Jackson & Messick, 1957; and Leavitt, Hax, & Roche, 1955). In these studies the investigators used reversed F scales written so that items in the reversed scales were presumably opposite in meaning from items in the original scale. Although all of these studies seem to indicate that a response set to acquiesce, independent of attitudinal content, plays some role in determining scores on the F scale, there has been disagreement on the relative importance of acquiescence and the role it plays in relation to authoritarianism.

In a comprehensive critique of this body of research, Christie et al., (1958) conclude that the lack of consistent positive correlation between the F scale and several independently constructed reversed F scales should not be attributed to a tendency to acquiesce. They state that the lack of consistent positive correlation is more a function of inadequate reversal so that reversed items may be assessing different rather than opposing content. Brown and Datta (1959) present evidence that seems to support Christie's contention. They found that correlates of intelligence remain significantly related to authoritarianism even with response set removed.

It appears as though the relative importance of acquiescence in F scale performance cannot be evaluated by using reversed scales until it is clear that reversed items are comparable to the original items in all respects but direction of wording. The purpose of the present study is to evaluate the adequacy of reversed F scales from the standpoint of equivalence and validity. Equivalence would require a correlation between reversed scales approaching their respective reliabilities as well as correlations with the original F scale in the same direction and of roughly the same magnitude. Validity demands significant correlations of the reversed F scales with an external criterion of authoritarian tendencies.

METHOD

Authoritarian scales

Of the five reversed F scales reported in the literature, the Bass (1955) and Jackson-Messick (1957) versions seem to represent the most widely varying approaches to item revision. Bass made relatively minor modifications of the content and structure of the original items while Jackson and Messick made considerably greater modifications in accordance with their desire to preserve only the style of the original statements.

Possible scores on the F and reversed F scales range from 1.00 to 7.00 with high scores on each scale indicating authoritarianism. Accordingly, a positive correlation between any two scales indicates consistency of response to content.

Criterion measure

The semantic differential consists of a set of concepts (e.g., *Jew*, *Homosexual*) which are rated upon a list of seven-step bipolar scales (e.g., Intelligent-Stupid). Twenty scales were developed on the basis of Osgood's (Osgood & Suci, 1955) analysis of "meaning" and his criteria of selection (Osgood, Suci, & Tannenbaum, 1957, pp. 76–78). To prevent the formation of position preferences, the polarity direction of the 20 scales was varied throughout the list with respect to favorability (e.g., Intelligent-Stupid was followed by Bad-Good).

A postulated tendency of the authoritarian personality is to regard group differences in "all-or-nothing" terms rather than in terms of gradations (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950, p. 228), and there is some empirical evidence to support this contention (Stagner & Osgood, 1946). Therefore, the authoritarian might be expected to justify differences by attributing unfavorable traits to outgroups in an all-or-none fashion. The outgroups judged in this study were *Jew*, *Homosexual*, *Atheist*, *Murderer*, *Prostitute*, and *Lunatic*. In addition to these socially controversial concepts, six noncontroversial concepts were used (*Waitress*, *Parrot*, *Postman*, *Cat*, *Fisherman*, and *Dog*) to assess any differences in scale-checking behavior as a function of concept controversiality. Accordingly, it should follow that *S*s scoring high on an authoritarian scale would have a greater tendency to make extreme judgments than low-scoring *S*s and this tendency should be greater with respect to controversial social concepts than with respect to noncontroversial concepts.

---

[1] This paper is based on a portion of a master's thesis submitted to the Department of Psychology of the State University of Iowa. Grateful acknowledgments are extended to Alfred B. Heilbrun, Jr., thesis advisor, and Harold P. Bechtoldt, who provided invaluable suggestions.

A measure of extremity of judgment on the semantic differential was obtained for each *S* by calculating the mean absolute deviation from the midpoint of the seven-point scales. These values were summed first over the 20 scales on each concept, then averaged over all six controversial concepts. The values range from .00 to 3.00, a score of 3.00 indicating an invariable use of the extreme positions. A separate mean deviation score was obtained over the six noncontroversial concepts.

Subjects

The 315 *S*s used in this investigation were selected from the undergraduate courses in introductory psychology at the State University of Iowa.

Procedure

The administration was divided into two phases. In the first phase, all *S*s were given the semantic differential. In addition, 62 of these *S*s were given Forms 40 and 45 of the F scale, 63 *S*s the Bass reversed scale (Scale B), and another 63 *S*s the Jackson-Messick reversed scale (Scale C).

The second phase, about one week later, involved the administration of the F scale, Scale B, and Scale C to *all* *S*s with order of presentation counterbalanced.

Thus, at the completion of the second experimental phase, all *S*s had been given a twelve-concept form of the semantic differential, the F scale, Scale B, and Scale C. In addition, test-retest data were available on the three authoritarian scales.

RESULTS

The product-moment correlations between the three F scale versions and between each F scale and the criterion are presented in Table 1. It can be seen that the two correlations between the reversed F scales and the F scale are both significantly greater than zero but in opposing directions. Furthermore, these correlations are significantly different from each other ($t = 1.99$, $p < .05$) and the relationship between the two reversed scales ($r = .35$) does not approach the reliabilities of either scale (Scale B = .61, Scale C = .59). Thus there is little support for assuming the reversed scales to be equivalent to the F scale in content or to be equivalent to each other.

TABLE 1

INTERCORRELATIONS BETWEEN SCORES ON THE F SCALE (F), BASS REVERSED F SCALE (B), JACKSON-MESSICK REVERSED F SCALE (C), AND EXTREMITY OF ATTITUDE RATINGS ON CONTROVERSIAL CONCEPTS (E)
($N = 315$)

Scale	F	B	C	E
F	(.67)	.18**	−.14*	.22**
B		(.61)	.35**	.09
C			(.59)	−.04

Note.—Parentheses indicate test-retest reliabilities.
* $p < .05$.
** $p < .005$.

TABLE 2

MEAN EXTREMITY RATINGS ON THE CONTROVERSIAL AND NONCONTROVERSIAL CONCEPTS OF HIGH, MEDIUM, AND LOW F SCALE SCORERS

Range	F scale scores		Controversial concepts	Noncontroversial concepts
	Cut-off points	N	Mean	Mean
Upper 20%	>4.09	57	1.45	1.36
Middle 60%	4.09–3.09	197	1.32	1.26
Lower 20%	<3.09	61	1.16	1.14
Over-all		315	1.32[a]	1.27[a]

[a] These over-all means differed at the .001 level of significance ($CR = 3.13$).

The relationships between the F scales and scale-checking behavior failed to provide any evidence that the reversed scales are measuring authoritarianism. However, the original F scale did correlate significantly with the criterion and the differences between the F scale-criterion correlation (.22) and this correlation for Scale B (.09) and Scale C (−.04) were both significant at the .05 level.

The relative ratings on controversial and noncontroversial concepts of high and low scorers on the F scale (Table 2) were evaluated by cutting the distribution of F scale scores at the 20th percentile and at the 80th percentile, providing comparisons of high scorers (upper 20%), medium scorers (middle 60%), and low scorers (bottom 20%). The difference between mean rating extremity scores for the two types of concepts was statistically significant for both the upper 20% ($t = 2.25$, $df = 56$, $p < .05$) and the middle 60% ($t = 3.00$, $df = 196$, $p < .01$) subgroups. The extremity scores for the lower 20% were approximately the same for the two groups of concepts. Thus, the high and medium F scale scorers tended to make more extreme judgments with respect to controversial concepts than with respect to noncontroversial concepts while low F scale scorers did not make differential judgments on the two types of concepts. Further, it was found that the upper 20% of *S*s on the F scale showed significantly higher mean scores on both the controversial concepts ($t = 3.63$, $df = 116$, $p < .001$) and noncontroversial concepts ($t = 2.44$, $df = 116$, $p < .02$), than did the lower 20%.

DISCUSSION

The findings of this study offer evidence that the Bass and Jackson-Messick reversed F scales are not equivalent measures in that their correlation ($r = .35$) falls considerably short of their restricted reliabilities and since they enter into differing relationships with the original F scale ($r = .18$ and $-.14$, respectively). There was also no evidence to support the notion that either reversed scale is a valid measure of authoritarian tendencies since neither correlated significantly with the external criterion of extremity of social judgments. The findings of nonequivalence of reversed scales does not answer which one of the two (or perhaps neither) represents an adequate reversal, since the extent of the original F scale-reversed F scale correlations is presumably confounded with acquiescence set and other form or content factors. It should encourage caution, however, in equating research findings based upon these two reversal scales at least. The failure of both reversed scales to correlate with the criterion certainly does not lend confidence to the research utility of either.

The F scale was found to be significantly related to scale-checking behavior on the semantic differential. The relatively low correlation between these two measures ($r = .22$) is tempered somewhat when restricting characteristics of the sample are considered. The college *Ss* employed were quite homogeneous in performance on both instruments (F scale, $SD = .58$; controversial concepts, $SD = .47$) and the mean F score (3.57) was lower than the presumed neutral point (4.00), indicating that few *S*s, if any, consistently agreed with the items.

In spite of the sample restrictions, the relative performances of high and low scorers on the F scale were in the expected direction. The extremity scores of *S*s in the upper 20% of the F scale distribution were significantly higher than those in the lower 20% with respect to both controversial and noncontroversial concepts, indicating that high scorers tended to judge both types of concepts more extremely than low scorers. The extremity scores for low scorers did not differ for these two types of concepts, while both high and medium F scale scorers obtained significantly higher extremity scores on the controversial social concepts than on the noncontroversial concepts. Thus, it may be concluded that in comparison with medium and high scoring *S*s, low scorers have a tendency to avoid extreme judgments on both types of concepts. These findings are consistent with descriptions of the nonauthoritarian typically found in the literature. For example, Adorno (Adorno et al., 1950, p. 778) points out that the easy-going low F scale scorer is characterized by an extreme reluctance to make decisions or to commit himself.

SUMMARY

Three versions of the F scale (original F scale and two reversed scales) were given to 315 college students along with a form of Osgood's (Osgood, Suci, & Tannenbaum, 1957) semantic differential. The semantic differential consisted of six socially controversial and six noncontroversial concepts which were judged against 20 evaluative scales. On the basis of postulated characteristics of the authoritarian personality, it was expected that *S*s scoring high on an authoritarian scale would have a greater tendency to make extreme judgments than low-scoring *S*s and that this tendency would be greater with respect to controversial social concepts than with respect to noncontroversial concepts.

The results indicated that F scale scores were positively related to the tendency to make extreme judgments on the semantic

differential in the manner hypothesized. On the other hand, no relationship was found between either of the reversed scales and the semantic differential. In addition, since the correlation between the reversed scales was considerably less than their reliabilities, it was concluded that these scales were not equivalent forms even though both were worded and scored in the opposite direction from the original F scale. The utility of these scales was discussed in the light of these findings.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BASS, B. M. Authoritarianism or acquiescence. *J. abnorm. soc. Psychol.*, 1955, **51,** 616–623.

BROWN, D. R., & DATTA, LOIS-ELLIN. Authoritarianism, verbal ability, and response set. *J. abnorm. soc. Psychol.*, 1959, **58,** 131–134.

CHAPMAN, L. J., & CAMPBELL, D. T. Response set in the F scale. *J. abnorm. soc. Psychol.*, 1957, **54,** 129–132.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56,** 143–159.

JACKSON, D. N., & MESSICK, S. J. A note on ethnocentrism and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54,** 132–134.

LEAVITT, H. J., HAX, H., & ROCHE, J. H. Authoritarianism and agreement with things authoritative. *J. Psychol.*, 1955, **40,** 215–221.

OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50,** 325–338.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. of Illinois Press, 1957.

STAGNER, R., & OSGOOD, C. E. Impact of war on a nationalistic frame of reference. *J. soc. Psychol.*, 1946, **24,** 187–215.

(Received March 23, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

CASE REPORTS

BRIEF SUCCESSFUL PSYCHOTHERAPY OF A CHRONIC PHOBIC REACTION[1]

C. SCOTT MOSS
State Hospital No. 1, Fulton, Missouri

This is a report of the brief successful treatment of a phobic reaction in which the patient's repressive defenses were penetrated, leading to a simple, convincing demonstration of the childhood learning of the phobic response and its subsequent ramifications. A symbolic fantasy procedure, tailored to the patient's unique personality dynamics, was then instituted in order to consolidate the obtained insights and further desensitize the patient to the fear-provoking stimuli. Primary reliance was placed on several innovations in hypnotherapeutic technique.

TREATMENT PROCEDURE[2]

The patient, aged 45, an art instructor, stated that she had recently seen the movie, "The Three Faces of Eve," and it finally had encouraged her to seek treatment for a life-long, intense fear of dogs. She was also motivated by the desire to present her daughter with a puppy on the occasion of her ninth birthday, two weeks hence. She requested that, as in the movie, hypnosis be used to uncover the source of her phobia. The patient made a generally favorable initial prognostic impression, seeming intelligent, sensitive, and flexible. The problem seemed relatively circumscribed, the integrating capacity of the ego intact, and her current environmental adjustment adequate. Her response to initial induction procedures was favorable, and it was agreed that treatment would be undertaken for a maximum of seven to eight sessions over a three-week period. The remainder of the session was spent in training her in rapid entry into the termination of hypnosis. At the conclusion of the hour, she asked that the attempt be made first to recapture the lost memory of "something important" that had occurred during the birth of a child ten years ago.

At the beginning of the second session, additional training was instituted to insure her cooperation with hypnotic procedures. Following her wish, she was regressed to the delivery scene in 1948. It was revealed that her baby had died during delivery (as had two earlier children), and she believed that she had forgotten some vital fact concerning its death. However, while she recalled many long-forgotten details, she was unable to recapture the specific memory, though she narrowed it down to something the physician told the nurse immediately following delivery. The following posthypnotic suggestion was given: that the sought-after memory would gradually enter awareness and that if she sincerely wanted to remember it, she could probably do so at the next session.

The patient reported at the third session that the intervening period had been fraught with anxiety because she had found herself preoccupied with memories from the era in question. She also believed that she had recaptured the memory for which she searched. It was the physician's remark, "The baby is dead; if we wait, she may go, too." Hypnosis was induced and the delivery scene revivified. She immediately "heard" the physician remark, "Its neck is broken." The patient identified this statement as the elusive verbalization. She remarked upon awakening that one immediate benefit of these sessions was that for the first time she and her husband had been able to talk freely about this painful period in their lives. At the conclusion of the session, she was given the posthypnotic instruction that the meaning of her lifelong fear of dogs would begin to manifest itself between sessions It was further suggested that if these memories, thoughts, or feelings were too disturbing, they would be communicated to her in night dreams.

In the fourth session the patient reported the following dream:

> I kept my dreams. I did not dream Friday night, and I did not sleep much Saturday night because Jane [her daughter] was ill, but I did have a dream. Do you want me to read it to you? This was Saturday night. It was about two little girls, toddlers. It was a very disturbing dream, incidentally. One child was playing with the other as if she were a doll. She was

---

[1] Appreciation should be extended to my graduate class in Psychotherapeutic Theory and Methods at the University of Missouri for their stimulating suggestions in discussion of this case: Phyllis Berger and Patricia Shallenberger, and James Hart, Fred Stern, Lynn Ourth, and Lyman Riegel.

[2] The entire therapy sequence was recorded and is available in transcript form for teaching and training purposes through arrangement with the author.

carrying her, putting her face next to hers, but the one that was being carried around was dead. I went to sleep worrying about Jane, so that might account for that. It seems as if the live one was my little girl, and I was distressed when she picked up the dead one because she did not realize she was dead. I asked the undertaker what sort of embalming fluid he used and whether it would hurt my little girl. My concern seemed to be for her. Then there was something about redressing the dead baby, and I was concerned about this. The undertaker or somebody was dressing it on a bed and I was very sad because it was so still, and then it seemed as if I was trying to get a box of a specific size, but I could not remember what the size was. I guess I needed the big box to put the little ones in. It seemed like I needed them for clothes or toys or something. And then it seemed as if I were at my sewing machine mending a badly torn dress. Suddenly I realized my child was dead and I cried a lot in my dream. Then it seemed there was only one child, the dead one, and it was mine. It was a very sad and depressing dream. I never had a dream like that before. It was horrible even when I thought of it in the daytime. And I didn't have anything to drink before I went to bed.

She also volunteered that, "Oddly enough when I woke from that horrible dream, I recalled my little sister's death, and I haven't thought of that for many, many years. I recall practically nothing about it. It is so long ago I just never think about that. . . . It must have been about 41 years ago, I was about four." Other associations were not immediately forthcoming, so hypnosis was induced, and it was suggested that she allow the full meaning of the dream to emerge.

For a ten-minute period, the patient sat in silent contemplation and then, at the prompting of the therapist, began to share a flood of memories which she apparently had repressed since childhood. The essence of the recovered content was that at age four, she had been playing in the backyard of their home with her younger sister when the family dog, Rover, knocked the little girl down, and she suffered a splinter wound in her cheek. Several days later, the little girl died, apparently of an infection, and the mother openly blamed the patient, accusing her of knocking her sister down.

The patient recalled the day of the funeral in some detail, particularly an incident which added to the traumatic impact of the whole situation.

Pt: . . . Then I go back in [the parlor] by myself. She looks so pretty lying there, so I think I will pick her up. She does not hug me like she usually does, though. I am holding her, and the dog is with me. He wants to see her, too. Oh, he tore her dress!

T: Pardon?

Pt: The dog, he tore her dress!

T: How did he do that?

Pt: He was jumping up on her and tore her dress, and Mother will not like that. The dog should not be in the house, but I do not think I let him in. I know I should put her down but I want to hold her.

T: What do you think and feel about her?

Pt: She is so pretty and so sweet. I just want to hold her.

T: What happens next?

Pt: My mother comes into the room. She is mad at me. She told me I should know better. Somebody took the baby away from me.

T: Did you want to give her up?

Pt: No, I wanted to hold her. It must be my Aunt Anna. She said I should not have torn her dress, but I didn't tear her dress.

T: [Simulating the Mother] Somebody did! Who did it?

Pt: I didn't do it. Rover did it; I didn't. He didn't mean to, he just wanted to see her. Why doesn't that place on her face go away?

Within a few days after the sister was buried, the patient's attitude toward her little dog changed markedly.

Pt: Now it seems like it is some days later. I am on the back porch by myself. I wish she would come play with me. Mother comes out the back door, and she finds me crying. I don't want her to see me crying.

T: You don't want her to know how you feel?

Pt: No. She thinks I hurt her, but I didn't.

T: Your mother blames you for what happened to your sister?

Pt: She thinks I pushed her over, but I didn't. I don't seem to like my dog anymore.

T: He still wants to be friendly, doesn't he?

Pt: I don't want to be friends.

T: You don't like your little dog?

Pt: No, I don't like my little dog anymore. I want my sister.

T: Your little dog took your sister away from you?

Pt: I don't like him. I wish that they would take *him* away.

T: You want him to go away and your little sister to come back?

Pt: Yes.

Prior to awakening the patient was given instructions to the effect that she was free to remember as much of this experience as she was prepared to accept. Upon awakening, she obviously recalled the entire episode.

T: Did you consciously recall your little sister's death?

Pt: I remembered very little about it. I remember she died. I heard them talk about it. The doctor in this little town didn't know what was the matter. They removed a splinter from her face. He thought it was an infection perhaps, but I didn't know what had happened.

T: Did you ever suspect that your mother had blamed you?

Pt: Oh, no, I didn't know Mother had blamed me. I can understand, though, some of my actions as a child. I always wanted to please everybody. I can remember that at school I would take blame for things that other kids did, and my mother would get real provoked with me because every week she had to go to the principal. If the teacher wanted to blame somebody, I was always glad to take the blame. As I grew older, I realized it was real stupid, but I guess I didn't want to be blamed for anything. I mean I didn't want anybody to be blamed for anything, so I took the blame. I was mixed up, I know that. I wanted people to like me as a child. I expect that Mother must have been terribly upset about me picking up the baby. Of course, if she believed I had caused her to fall, she would have blamed me.

At the conclusion of the session, a posthypnotic suggestion was given that in the interim, the meaning of this experience and its total effect upon her life, past and present, would begin to make itself evident to her. The patient also expressed an intention to write her mother and discuss the event that therapy had uncovered.

At the beginning of the fifth session, the patient volunteered, "If there is any question about our making progress, we are." She stated that for the first time in her (adult) life, she had petted a dog. She also reported a reassuring dream in which she was attacked by a large, vicious dog but managed to calm him.

In the three days between sessions, the patient had experienced the compulsive recall of a series of memories beginning in the recent past and extending back to early childhood. In the waking state and under hypnosis, these experiences were explored in detail. Through them ran a theme of feeling misunderstood and mistreated by her family, strong resentment and associated guilt, and a continuous struggle for independence and self-realization.

One memory concerned an event at age eight involving a puppy her parents gave her in hopes she would overcome her fear. She reported in detail an episode in which she feared the dog might attack another younger sister, and how immediately afterward, it had a fit and died. This experience apparently reinforced the potency of the origial trauma. Careful exploration revealed that the patient seemed to have no conscious knowledge of the reason underlying her intense fear of dogs, even at this age. The patient subsequently was regressed to successive age levels, but no evidence whatsoever could be found to indicate that she had ever experienced a physical attack by a dog which would account for her fear. This indirectly testified to the validity of the early experience recalled in the previous session in establishing her phobia. The patient, incidentally, had been able to verify many of her recently recovered memories pertaining to the past twenty years by reference to a meticulous, daily family diary and file of all letters they had ever exchanged, kept by her husband. A posthypnotic suggestion again was given that the full meaning and effect of this early experience would continue to manifest itself.

In the sixth hour, the patient appeared more relaxed than at any time since the first session. She verbalized a number of new found insights; for example, she now understood why she had always had the need to be liked and the fear of offending others, the basis for her deep seated ambivalence toward her mother, and the reason underlying her discomfort in disciplining her daughter. Another insight pertained to her revulsion at touching dogs: the lively or frisky feeling of the dog brought back, through contrast, the limp feeling of her little sister experienced at the time of her funeral.

During the session the technique of age regression was again employed, and the period between her early traumatic experience and the one at age eight, uncovered in the previous session, was explored. So far as could be determined, conscious memory of the earlier event ceased to exist even prior to age five. To the end of determining what there had been in the movie, "Three Faces of Eve," that finally had prompted her to seek professional help, the patient was instructed to project on an imagined movie screen the scenes from the movie which had, consciously and unconsciously, the greatest personal meaning for her (Moss, 1957; 1958). A fairly strong identification with the movie heroine was revealed, based on the similarity of childhood experiences relative to death and the resulting feeling that no one, parents or spouse, seemed to understand the consequent emotional disturbance. The movie also increased her optimism that she might be helped, especially through hypnosis.

A major therapeutic tactic designed to facilitate the relearning process was introduced near the conclusion of the session. The patient was given the posthypnotic suggestion that between sessions, whenever she was alone and desired to re-establish her acquaintance with Rover, the dog of her early childhood, she could experience a positive visual hallucination of him. It was emphasized that the dog *symbolized* her disturbing history, and this fantasy relationship provided her with the opportunity to make peace with her past. It was further emphasized that she had complete control of the relationship and could "allow the dog as close as [she] desired or keep him as distant as [her] fear dictated."

In the seventh session the patient reported that she had written to her mother concerning the recovered disturbing memories and that although the mother had confirmed the existence of a small, black dog, she insisted that the patient's fear was the result of prenatal conditioning. This seemed quite inconsistent to the patient since, from other sources, she knew that they had had several family dogs prior to the time when she was four and had never had another except for the puppy when she was eight.

The patient reported increased feelings of comfort in her general relationships with dogs. She had been able to experience Rover's hallucinated presence quite vividly and also had succeeded to some extent in making friends with him. She laughingly referred to him as "my little black Harvey." At this juncture, the patient agreed that sufficient progress had occurred to warrant a month's vacation from therapy; however, she requested that she be allowed to keep Rover in order to renew further their acquaintance. She reported that her feelings toward him were no longer those of fear or even sadness, but instead "like two old friends who have quarreled and don't quite know how to make up."

She was given the posthypnotic suggestion that she continue to work through her problem in the following form:

> Of course, it will not be possible for you to keep Rover indefinitely. As you pointed out today, it would be so nice to have Rover really there. I think you can have Rover *really there* in a sense that down through the years Rover has always been there in every dog that you met, particularly in little, frisky dogs. Each dog has been Rover as you looked at him with your little-girl eyes. If you work hard at becoming friends with Rover, your fear will leave you and you will quite naturally develop a greater acceptance of all dogs As time goes on you will develop the idea of having a real Rover for you to enjoy and especially for Jane, who needs and wants a Rover of her own.

A relatively brief session was held one month later. The patient reported that five days after the previous session she had gone skating with her daughter and met several dogs. "It seemed very natural to reach down and pet them. There was no fear of doing it. They did not *feel* like dogs used to feel. I heard Jane yell, 'Oh, look, Mommy used to be afraid of dogs and isn't anymore,' and then it dawned on me that I really wasn't. It seemed that right after that I lost Rover." Laughingly she inquired, "You don't have him around here anymore, do you?" She stated that the family was now actively looking for a dog. She also presented the therapist with an unsolicited fifty-nine page narration concerning her daily feelings and reactions during treatment. Three weeks later, the patient phoned to report that she had purchased a black cocker spaniel puppy and had named it Happy Siesta.

Almost one year to the day later, a follow-up interview was held with the patient. She began the session with the question, "Suppose the event I recalled wasn't true: could the mind make up something like that?" Discussion revealed other evidence of some residual resistance against a complete emotional acceptance of her childhood trauma; for instance, some elements of the experience were now incompletely remembered, and she described the memory as "not quite real." She also reported that while her mother continued to deny that the event had ever happened, a brother had confirmed the existence of a family dog named Rover and also the fact that she and her sister had been left momentarily alone at the time of the fatal accident because the two brothers, who had been carrying in wood, had gone to witness a fire at a neighbor's. Her younger sister also recalled that the patient had anxiously protected her in the presence of any and all dogs.[3]

It was explained to the patient that it was relatively unimportant whether the memory of her childhood experience was correct in every detail, that the important factor was the effect of the therapeutic experience on her present adjustment. She reaffirmed that therapy had somewhat altered her relationship with her mother, daughter, and husband. The most evident effect of psychotherapy, however, continued to be a marked change in the patient's attitude toward dogs. "All my life the touch of one gave me a feeling of something horrible. I'm still surprised that I feel so differently now."

DISCUSSION

This case provides a clear demonstration of the use of a psychotherapeutic procedure in modification of a long standing learning process. Adjunctive hypnotic techniques were of especial value in rapidly uncovering and revivifying the childhood pathogenic experience, allowing the adult ego to grapple with it in a new attempt at mastery.

---

[3] There was considerable circumstantial evidence to support the occurrence of a traumatic experience of the nature reported by the patient. Most convincing, possibly, is a congruence to the known dynamic structure basic to many phobic reactions; that is, the phobia dates from a fear-provoking episode in childhood, usually involving a severe disturbance in the parent-child relationship. Associated anxiety, shame, and guilt result in suppression and eventual repression of the frightening experience, leaving only the residual phobic symptom.

Relearning was further facilitated by reinstatement on a symbolic fantasy level of the patient's relationship with the original object of her phobia and with sufficient vividness to make it a realistic and immediate emotional experience. Hypnosis was also of value in reinforcing the patient's motivational level and at times in providing an almost compulsive re-experiencing of long-forgotten memories. In this manner, substantial progress was facilitated between sessions and following the termination of formal sessions, apparently significantly shortening the period of actual treatment. Success in this direct relearning procedure was naturally dependent on the assessed capacity of the patient's ego to assimilate repressed infantile experiences into conscious awareness.

The foundation for the patient's neurosis appears to have been a lack of understanding, acceptance, and emotional warmth in relation to her parents, particularly the mother. The traumatic incident in which she was unjustly accused of her sister's death focused her feelings of being mistreated, misunderstood, and rejected onto dogs. Specifically, the physical sensation of her sister's cold, still body became associated with its opposite, the warmth and liveliness of the real culprit, her dog. In a sense, dogs thus became an idiosyncratic symbol of the disturbed relationship with her mother.

In spite of her use of the dog as an object for projection and her protestations of innocence, continued accusation and rejection by her mother seem to have generated some guilt in the patient. This guilt appeared in her eagerness for punishment as a child, which may be variously interpreted as direct atonement, an assertion of innocence by being unjustly punished (a repetition of the circumstances surrounding her sister's death), a technique for forcing her mother to support her innocence by having continuously to get her out of trouble, or all of these. Some feelings of guilt were also apparent in the deaths of her children, and it appeared very important for her to prove to her husband that she was not to blame.

One important consequence seems to have been the patient's inability to express directly her feelings of hostility and resentment. A hypothesis that had to be considered during treatment was that she originally had projected these feelings onto her dog and that her phobia protected against their expression. Her desire to present her daughter with a puppy (dependent on eradication of the "protective" symptom) might then have negative implications. Fortunately, the year witnessed some improvement in the patient's relationship with her daughter with no evidence of an increased display of hostility.

The approach employed encouraged the patient to re-evaluate the influence of her childhood experiences in terms of present day understanding and to effect a fuller readjustment, not only to the original conflict, but to its ramifications in her present day life. No claim is made for achievement of a complete cure in the sense that the patient achieved full readjustment to all aspects of the widely generalized effect of the original conflict. Neither should the impression be conveyed that hypnotherapeutic techniques afforded a panacea per se. The patient described her youth as characterized by "a fear of people, new situations, many things," and her marriage as having had a highly therapeutic effect. The death of a beloved older sister when the patient was aged 27 helped the patient partially to overcome her fear of death. Factors such as these contributed to a substantial degree of "spontaneous recovery" and established a "readiness" to circumvent the repressive defense upon which the hypnotherapeutic procedure then capitalized.

REFERENCES

Moss, C. S. A forced hypnoprojective fantasy used in the resolution of pseudoepileptic seizures. *J. clin. exp. Hyp.*, 1957, **5**, 59–66.

Moss, C. S. Dream symbols as disguises. *Etc.: J. gen. Sem.*, 1958, **14**, 267–273.

(Received February 24, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

AN INTERVIEW WITH A MUTE CATATONIC

JOSEPH LYONS

Veterans Administration Hospital, Lexington, Kentucky

Of the classic symptoms of catatonia, muteness captures the interest of the clinician to a greater extent than does immobility or even *flexibilitas cerea*. Whether this is due to a recognition of the importance of language as a medium for social relations, or simply because of the high value that members of a "talking" profession place on speech, is not easy to say. In any case, a situation in which a mute patient is persuaded to talk promises to be both dramatic and revealing. In addition, as the following interview shows, it has much to teach us about the world of the schizophrenic, even when the patient offers little that seems to be explicitly informative or introspective.

The patient, Evan Lear, aged 29, was admitted to the hospital in a catatonic episode of acute, sudden onset. Seen in Admission Staff conference on the same day, he was mute and somewhat disheveled, responding passively when led into the room but resuming his tense, rigid attitude when seated. During the interview his facial expression varied from a flat, silly smile to obvious fear. Examined a few hours later under thiopental sodium (pentothal, Abbott), he said that he did not talk because "one person should not do all the talking" and because he did not want to "tell tales out of school." Although he was evasive and noncommittal, there was some evidence of paranoid thinking.

The next day he was given amobarbital (amytal sodium, Lilly), and while under the influence of the drug he ate voraciously and commented spontaneously about how hungry he was. During the following week, there was an unsuccessful attempt to administer psychological tests. In a second attempt, using amobarbital, he produced a Rorschach protocol and some drawings. These were interpreted as showing a long-standing schizophrenic condition with paranoid features in a man of average intelligence. The Rorschach responses were almost all Wholes, with good form, no color, and only one Human percept. On Card VII he saw "a woman broken in two," and to this he commented in the inquiry, "Looks like they've drawed a little too close together."

At the Medical Staff two weeks after his admission, it was reported that although he had been feeding himself with a spoon, he remained mute on the ward. He usually displayed a silly smile but rarely looked around him and had no interpersonal contacts. Occasionally he showed other symptoms—waxy flexibility to a mild degree, manneristic moving of his hands, and rapid blinking of his eyes. Except for a few impulsive outbursts of aggression, he remained essentially unchanged for the next month, while permission to begin electric shock treatment was sought from his father. Meanwhile he had been studied intensively, by means of tape recording and motion picture photography, in the Research Laboratory of the hospital. Approximately nine weeks after his admission, the interview which is reproduced here was obtained as part of a continuing study of catatonic expression. Two psychiatrists and a research psychologist (identified in the text, respectively, as Drs. A, B, and C) collaborated in an attempt to have the patient reveal his experiences in the stuporous condition.[1]

INTERVIEW

Dr. A.: We hope we can help you. —To understand you better, we have given you some drugs which will make it easier for you to talk. Perhaps then you can tell us something about yourself, and about the way you look at the world. —That isn't painful, huh? —Do you have a little bit of a strange taste? —You have been with us now for quite a long time, and in all these days you haven't told us much. Do you think you could do it today? —If you don't want to talk, you could nod. Could you do that much? —You may get tired. Do you feel tired? Huh? Do you feel tired? Couldn't you give me a sign, just a yes or no? —Feeling tired? Huh? How is it? Feeling sleepy? —Say, do you know that this is a hospital here where we are? Do you know that? That we are doctors? Do you know that? This hospital is at Lexington, Kentucky. Do you know that? You haven't given me any sign whether you know it or not. That is not much to say, yes or no. Huh? Are you drowsy? Say, do you feel drowsy, sleepy? —Would you like to lie down? Would you like to lie down?

---

[1] For the interview, the patient was prepared with an intravenous injection of 3.75 grains of amobarbital. The entire session, which lasted approximately 50 min., was tape recorded. It is given here with no editing except for necessary punctuation; the reader will recognize that pseudonyms have been used for the real names of persons and places.

It should be kept in mind that the interview actually proceeded quite slowly and that the impression which may be gained in reading, that the interviewers were pressing or badgering the patient, results from the necessary compression of the spoken material on the printed page. The dashes are meant to represent pauses; however, no effort has been made to show pauses between speakers. Actions by the patient are given in parentheses; the bracketed and indented portions are comments by the writer.

Dr. B.: We know you can talk. Suppose you tell us your name. —What is your name? Huh? What is your name? —Is your name Smith?

Dr. A.: Who are you? Now tell us, who are you? —Then what are you? Okay, who are you? What are you? Tell us, please. Are you one of us, or are you somewhat different? —Getting sleepy, huh? Getting sleepy? Keep awake, can you? Can you keep awake, huh?

Dr. B.: Supposing you tell us your name. What is your name, huh? What is your name?

Dr. A.: Could you be good enough and tell us who you are?

Dr. B.: What is your name? Come on, now, tell us. Carl? John? Tell us your name—what is it?

Dr. A.: Do you hear us? Don't be so proud you cannot give a sign. Give us a sign whether you hear us or not, huh? Huh? Hello. —You would rather go to sleep than to speak, huh? Can you open your eyes?

Dr. B.: Open your eyes. Look up here. Come on, look up here. Look up, now.

Dr. A.: Do you want some food, something to eat? —Who are you? Who are you? Come, let us know. What are you?

Dr. B.: Look up here. Who are you? Who are you? Open your eyes, now, and look up here. Come on. Open your eyes—wider, wider. Open them real wide. Now look up here. Look up. Open your eyes and look up here. Look up, now. Who are you? Tell us who you are. Who are you?

Dr. A.: Perhaps you could write it down if you had a paper and pencil.

Dr. B.: Who are you? What's your name? What is your name? What's your name? What's your name? Tell us who you are. Tell us who you are. Look up, now, open your eyes. Open—open your eyes and look up. Who are you? (Patient is given paper and pencil, which he takes.)

Dr. A.: Would you write it down? Would you write down who you are? I will hold it for you. And what are you? What are you? Write it down. (Patient writes his full name.)

Dr. A.: That says who you are, but we would also like to know what you are. Is that all you can do?

Dr. B.: What does that say? What does this say?

Dr. A.: Could you write down how old you are. (Patient writes some numbers.) Thirty-one years old?

Dr. B.: What is that? Serial number. . . .

Dr. A.: Yes.

Dr. B.: Was that Army or Navy? Was that Army or Navy? —Is your name Evan? Is your name Evan? Is that your first name? Is it Evan? Evan? Is that right? (Patient nods.) It is Evan. And were you in the Army or Navy?

Dr. A.: Write it down. (Patient writes "Army.")

Dr. B.: Now can you tell me what that says.

[Of interest in the interview thus far is the eagerness of the doctors to have the patient say something, no matter what it is. Now that he is about to utter his first words in months, one almost expects that, like the Sphinx, he will reveal a profound secret that will not be an anticlimax to his profound silence. But notice the interchange that follows.]

Pt.: Ar—my.

Dr. B.: How long were you in the Army? What's that? Huh?

Pt.: Two.

Dr. B.: Two what? Two years? Is that right, two years? Is that the right time you were in the Army, huh?

Pt.: Uh-huh.

Dr. B.: And where do you live?

Dr. A.: Come, write it down. (Patient writes an address.)

Dr. B.: I can't read that too well. Can you tell me what it says?

Pt.: Care of Edgar Lear, R.F.D. 2, Esquirol, Ohio.

Dr. B.: Is that your home?

Pt.: Yes sir.

Dr. B.: And how long have you been there, Evan?

Pt.: I have been there all of my life.

Dr. B.: And do you know where you are now?

Pt.: Approximately.

Dr. B.: Approximately, where? What kind of a place is this?

Pt.: A V.A. hospital.

Dr. B.: That is right. Now, do you know what state it is in?

Pt.: Kentucky.

Dr. B.: Do you know what town this is?

Pt.: This is Lexington.

Dr. B.: And how long have you been here now?

Pt.: How long?

Dr. B.: Uh-huh.

Pt.: Two months and ten days.

Dr. B.: Two months and ten days. Do you have any idea why you are in the hospital?

[At this point the patient becomes silent, as can be seen from the passage that follows. In order to understand this, we have to consider the kind of question that produced the silence. Note that it differs from the preceding ones in that it is not a question of fact. The answer to it requires that the patient consider his present situation with some degree of reflective awareness, that he adopt toward himself an attitude different from the attitude one has toward the objective, public data of one's life. This he apparently cannot or will not do. The first hint of his failing in this regard, in fact, was given when, to the first question about where he was, he changed his previously precise answers to the slightly evasive and imprecise word, "Approximately."]

Dr. A.: You have been talking so little—why? Is some one forbidding you to talk?

Dr. B.: Do you have any idea, Evan, why you are here? Is there something the matter with you? Huh? —Are you sick? Evan, are you sick? —You haven't answered me. Are you sick, Evan? —We think you are sick. Are we wrong? Or are we right? Huh? What would you say? —Evan, what would you say about that? —How do you feel? Huh? Evan, do you feel all right? Come on, now, you are not that sleepy. How do you feel? —Huh? Do you feel all right? Evan.

Dr. A.: Do you have any orders from God to keep silent? Does God order you not to talk?

Dr. B.: Evan, look up here a minute. How do you feel? —How do you feel? Talk to me, how do you feel?

Do you feel well? Huh? Do you feel well? Does anything hurt you? Does it?

Dr. A.: Would you like to write it down—how do you feel?

Dr. B.: Evan, look up here, now. Tell us how you feel. Does anything hurt you? Huh? —Nothing hurts you. Do you feel good? Do you feel good or bad? Huh? —Come on, now, you know whether you feel good or bad. Tell us. Which is it? Evan, which is it? Huh? —Come on, spit the word out—which is it, good or bad? —Aren't you going to tell us? Or don't you know, huh? Evan—

Dr. A.: Do you want to write it down? You could write "good" or "bad." Could you?

Dr. B.: Why don't you look at (Dr. A.) when he talks to you? Come on, now, and look at him. He is not such a bad looking guy. Turn your head around here and take a look at him. Go on. Look up at him. Evan.

Dr. A.: Do you have a secret which you don't want to reveal? That you don't want to reveal? Evan, do you have a secret? No secret. (Patient is given a drink of water, which he accepts.)

Dr. B.: How are you feeling now, better? A little drunk? Huh?

Dr. A.: Could you talk a little bit more?

Dr. B.: Would you like to have another drink? Huh? What do you say? —You look like you are feeling pretty good now. Do you? Do you? Huh? —You look like you are enjoying this—are you? Does this make you feel pretty good, Evan? Evan, what do you say? Are you enjoying this? —I have some more where that came from. Would you like to have some more? Huh? How about it? Do you want some more? —What do you say, Evan? Do you want to quit it or do you want some more? Would you like to continue?

Dr. A.: Do you have a brother or a sister? Have a brother or sister?

Dr. B.: Evan, Dr. A. asked you a question.

Pt.: Three.

[The question that brings the patient back into the conversation is a straightforward question of fact.]

Dr. B.: What?

Dr. A.: Three sisters?

Dr. B.: I can't make much out of that. You will have to tell us. Brothers, sisters, both?

Pt.: Two.

Dr. B.: Two what? Two what, Evan?

Pt.: Brothers.

Dr. B.: Do you have any sisters?

Pt.: One.

Dr. A.: I think he said four.

Dr. B.: How many, Evan?

Pt.: One sister.

Dr. B.: Is your sister older than you are, or younger?

Pt.: Younger.

Dr. B.: Is she married?

Pt.: Yes.

Dr. B.: How about you, are you married? —Have you ever been married?

[The patient does not answer this pair of questions. He is selective, even in regard to questions of fact.]

Why not? Don't you like the girls? Huh? Or haven't you found one you would like to know? —Does your married sister have any children?

Pt.: One.

Dr. B.: Then that makes you an uncle, doesn't it? —And how about your brothers—are they married? Are they younger than you or older?

Pt.: They are younger.

Dr. B.: Then you are the oldest in the family—is that right? You are the big brother?

Pt.: Uh-huh.

[Again, as the questioner touches on an area about which the patient might have some regard for its significance in his life history, Evan shies away. It should be noted, however, that it is not simply the biographical or "personal" questions at which he balks. He will answer questions about himself. The distinction is that he will respond only to questions which do not call upon him to express some measure of reflective understanding of his position as a meaningful point in his personal history. Thus, the result of the interview so far is that some information has been obtained which one might get more easily, and with greater confidence in its accuracy, from the patient's case folder.]

Dr. B.: And do they like their big brother? Huh?

Pt.: I couldn't tell you.

[At this point, for the first time, the patient does not fall silent at a question of personal significance but simply pushes it away. We will see below that he has other methods at his disposal for dealing with such pertinent questions.]

Dr. B.: Oh, I bet you have some ideas about that. You really couldn't say? —Are your folks living, Evan?

Pt.: Yep.

Dr. B.: How old is your dad?

Pt.: I figure around fifty-four.

[This is an excellent, well-phrased answer, only one of many well-spoken sentences which occur throughout the interview. Could one guess, on the basis of this answer, that the patient is only a few moments or a few words away from one of the most striking of schizophrenic symptoms? Evidence of this kind makes it difficult to maintain that the patient is the helpless victim of an overwhelming emotional state or of a pervasive disorder of thinking. The interchange that follows is particularly illuminating in this respect.]

Dr. B.: What is his name?

Pt.: Edgar P. Lear.

Dr. B.: And your mother, how old is she?

Pt.: How old she is?

Dr. B.: Probably about the same age your father is?

Pt.: Close.

Dr. B.: And what is her name?

Pt.: Cordelia Regan.

Dr. B.: And they live in Esquirol, Ohio?

Pt.: No, they have been separated.

Dr. B.: And who have you been living with?

Pt.: My dad.

Dr. B.: And where does your mother live?

Pt.: Gone to Goneril, Ohio.

Dr. B.: Ohio?

Pt.: Yes.

Dr. B.: Is that very far from Esquirol?

Pt.: About twenty miles—twenty-five.

Dr. B.: That is not so far. Do you ever visit your mother, see her? Does she ever come to see you?

Pt.: I see her every once in a while.

Dr. B.: Is she re-married again?

Pt.: Yep, a Regan.

Dr. B.: She is married to somebody by the name of Regan. And who is there at home with your dad?

Pt.: There wasn't nobody when I left.

Dr. B.: The others were all away?

Pt.: Uh-huh.

Dr. B.: Just you and your dad live there together?

Pt.: Yep.

Dr. B.: What sort of work did you do?

Pt.: Made poultry coops.

Dr. B.: Poultry coops. Are you a carpenter—that sort of thing?

Pt.: No, I'm not a good carpenter.

Dr. B.: Well, how do you make poultry coops—what do you use? Lumber and chicken wire, is that what you mean?

Pt.: I used any kind of wood you can get hold of.

Dr. B.: Uh-huh. And do they raise a lot of chickens around that part of the country?

Pt.: Quite a few.

Dr. B.: And these coops you made, you sold—they weren't just for your own use? You made coops for others too?

Pt.: They were for use to put poultry in.

[Clearly, this patient is capable of a kind of tart dignity which sometimes even makes the questioner seem a little foolish.]

Dr. B.: And how was business—was it pretty good?

Pt.: Yes, it was pretty good.

Dr. B.: Were you working before you came to the hospital?

Pt.: Yep.

Dr. B.: How long hadn't you been working before you came to the hospital?

Pt.: What was that question again?

[Here, if anywhere in the interview, is evidence that the patient is in contact, even to the extent of asking for clarification of a question.]

Weren't working?

Dr. B.: Yes. Did you work before you came in, that is what I mean. Or had you not been working for a couple of weeks? Do you remember now? Huh? —When you came in, Evan, you weren't talking to us. Did you stop talking a long time before you came to the hospital?

[At long last we come to the topic which is the real purpose of the interview. What will the patient say when he is finally confronted with the issue of his muteness?]

Pt.: I never did talk too much.

Dr. B.: No, but you talked some.

Pt.: Yep, a little.

Dr. B.: And did you stop talking after you got here or before you came in?

Pt.: Before I came here.

Dr. B.: Was it a long time before or just a couple of days? Huh? Do you remember now how long it was that you had stopped talking before you came here?

Pt.: I don't know just how long.

Dr. B.: Do you have any idea why you stopped talking? Huh? —What made you stop talking? Huh? —Did anything make you stop talking, or did you just decide to quit? Huh? Which was it? Evan, huh? —Did you just decide to quit talking or did something make you stop? Huh? Evan—why did you stop talking? Huh? —Evan, you haven't answered my question. Why did you stop talking? Huh? —Aren't you going to answer that question? Huh? What do you say?

Pt.: I couldn't say yes and I couldn't say unh-uh.

Dr. B.: You mean you don't know why you stopped talking?

Pt.: No.

Dr. B.: Did you get some message from some place that wanted you to stop talking?

Pt.: No.

Dr. B.: Isn't it kind of hard to get along with people and not talk to them?

Pt.: Not always.

Dr. B.: Do you think it is easier by not talking?

Pt.: I wouldn't say that.

Dr. B.: Well, you have been here, you have told me, three months and ten days?

Pt.: Approximately.

Dr. B.: And you have hardly said a word. You know, we would like to know how you feel and what is bothering you—and maybe we can help you. Would you like to get well and go home? Huh? —Would you like to get out of the hospital, Evan? Huh? Or haven't you given it any thought? Huh?

[At this point the patient has become silent again. What have the interviewers achieved thus far? The patient has given some information by answering questions of fact; he has told us that one of the common explanations for his silence, that he received a "message" to this effect, is not true for him; on matters of opinion he remains noncommittal; and on questions concerning cause or significance, he simply does not talk at all. Thus, he leaves us in doubt at the very point where we most desire an answer—although this finding may itself be something from which we can learn.]

Dr. A.: Do you like it here? Huh? Do you like it here?

Dr. B.: You can say yes or no to that one, can't you? Huh? Do you or don't you?

Dr. A.: Just say no if you feel that way.

Dr. B.: Do you like it here in the hospital, Evan? Huh? How have they been treating you here? Have they been treating you all right? Huh? —What do you say? Has everybody been treating you all right? Huh?

Dr. A.: Would you like to talk if you could? Huh?

Dr. B.: Evan, why don't you answer (Dr. A's) questions? Would you like to talk? Huh? —Is it hard for you to talk, Evan? Does it hurt you to talk, or does something happen to you when you talk—that you would rather not, huh? Is that what's wrong? Evan, you know you have never told us why you don't want to talk. Why is it? Huh? Are you afraid talking will get you into some sort of trouble? Or that you might tell somebody a secret you don't want to? Huh? Evan, open your eyes. —You have never told us why you won't talk. Suppose you tell us. —Open your eyes. Tell us why you won't talk.

[It is ironic that the interviewers are still making these requests after the patient has indeed "talked."]

Dr. A.: Do you think we are not good enough to talk with you—that we are not good enough? Huh?

Dr. B.: Evan, what do you say? Would you like us to stop bothering you? Huh? —Are we bothering you, Evan? (Patient shakes his head.)

[He may not be talking, but he is not out of contact.]

Not bothering you. —Then why don't you talk? Huh? Tell us that? Why don't you talk? Evan. Why don't you talk to us?

Pt.: There ain't much to say, is there?

[This is an astounding statement. What are we to make of it? Is it possible that the patient has heard and understood the coaxing and cajoling to have him talk, that he is aware of how important his talking is to his doctors, but that he simply disagrees with them that he ought to talk? This is one possible explanation of his remark, but it seems highly improbable. On the other hand, if we accept his statement as describing how things appear to him—as I think we must—then we may have to make some adjustment in our views concerning his "conflicts." Rather than saying that his silence is a determinate result of the psychic forces produced by his conflicts, we might propose that it is the other way around; that he maintains his silence in order to avoid a conflict to which talking might lead; that the decision to keep silent precedes, rather than follows, the presence or absence of conflict. This would be consistent, too, with the evidence that he is unerring in his selection of the kinds of questions that he answers.]

Dr. B.: Well, sometimes there is. —We could get along much better if you would talk to us. —Are you afraid of saying something that you shouldn't? Huh? Evan, are you afraid that you might say something which you shouldn't? —Doesn't it seem rather strange to you that a person just refuses to talk for three months? Now, you can say good morning. —Isn't that kind of peculiar, Evan? Huh? —Isn't it? —Tell me—doesn't that seem strange to you? Huh? —Doesn't that seem peculiar, Evan? —What do you think about that? —What would you think about someone who didn't talk for three months? Huh?

Dr. A.: Well, he feels he won't do it today, either.

Dr. B.: Oh, sure he will. —Evan, what do you think about somebody who won't even talk to anybody? Huh? Who wouldn't talk for three months? Wouldn't you think there was something wrong with them? Huh? —Did you have an argument with your father before you came into the hospital? Huh?

Pt.: No.

[Again the patient is perfectly "willing" to re-enter the conversation, but only on very specific terms. The questions must be of the type that usually appear on a face sheet or an application for employment—questions of fact, the content and accuracy of whose answers could be a matter of public record. The result of the patient's "decision" for muteness would seem to be this, that as far as others are able to know or discover, he is now completely a "public" creature; no longer a unique individual, with a private life which he hides or reveals, shares or keeps to himself, expresses or conceals, but something which only looks like a person but has no more life than a case file or a set of vital statistics.]

Dr. B.: You didn't get into any words with him? (Patient shakes his head.) Then why are you here? —Do you think you should be in the hospital? Huh? —Evan? What do you think about it? Should you be in the hospital? Huh? —What do you think about that, Evan? —Have you ever seen this gentleman before? Do you know his name? —It is (Dr. C.)

Dr. C.: We are old friends, aren't we, Evan? Huh? —Do you remember playing ball over there? —You drew me a picture once. What did you draw, Evan? Huh? —Do you remember you drew a nice picture, a nice big picture? What was it?

Pt.: A little pyramid.

Dr. C.: You very carefully drew a pyramid. Then you drew something on top of it—what was that?

Pt.: A little star on top of them.

Dr. C.: You were afraid the stars were going to fall off, weren't you? —Do you know what else we talked about over there? —You remember the bullet wound, don't you? —Where were you wounded? —Where were you when you got this bullet wound? Were you overseas? —Is it on your right foot or your left foot? —Oh, I remember now, it is on your right heel. —What year was that?

Pt.: 195—1945.

Dr. C.: You got out of the army right after that?

Pt.: Uh-huh.

Dr. C.: Were you a good soldier? —And wasn't that a funny question?

Pt.: I will let somebody else be the judge.

[Again the patient demonstrates a dignity which is rather admirable. Although he is committed not to be committed to anything as a person, some individual characteristics still shine through.]

Dr. C.: Were you a Private? Pfc?

Pt.: P-v-t.

Dr. C.: Were you a rifleman?

Pt.: Uh-huh.

Dr. C.: Where were you—France? Germany?

Pt.: I was in the Ruhr pocket.

Dr. C.: You were there, huh? How did you feel about getting out? Were you tickled? —That is a little tricky question, too.

Pt.: That is something I don't have much to say about.

Dr. C.: Say, when you say you can't say yes or no as to why you don't talk, is that because you don't know or because it is so complicated you can't explain? You know, it could be either way. It could be that you don't know or it could mean that it could be complicated or it is too personal—either way. What do you say about that? Huh? —Do you like ball—baseball? —Who won the World Series? —Do you remember the World Series? I don't know whether you were here. How about it? How far did you go to school?

Pt.: I didn't quite finish the second year in high school.

Dr. C.: How old were you, do you remember?

Pt.: Nineteen.

Dr. C.: You were the oldest child. Why did you quit school?

Pt.: Had to go to the Army.

Dr. C.: Did you say nineteen? The second year high school?

Pt.: Yep.

Dr. C.: Had you gone to school regularly?

Pt.: Every year.

Dr. C.: How far did you live from your school?

Pt.: Six miles.

Dr. C.: Did you ride the school bus? Consolidated school?

Pt.: I don't know whether you would call it a consolidated school or not.

Dr. C.: Did you go on the school bus, walk, or what?

Pt.: Rode the school bus.

Dr. C.: Do you know how many students there were in the school—would you guess? Was it a big school?

Pt.: The whole school?

Dr. C.: Yes.

Pt.: A hundred and fifty.

Dr. C.: It wasn't too small. You liked school and just had to quit to go to the Army?

Pt.: Uh-huh.

Dr. C.: Were you drafted?

Pt.: Yep.

Dr. C.: This effect of the sodium amytal—it makes you feel a little sleepy, of course, but how about moving your hand? Can you move it as well as you used to? Let us see your grip here. How hard can you squeeze my hand? —Not too hard. —Can you—going back—can you say why you can't talk, whether it is too complicated, you don't know, or because it is so personal you just don't want to talk about it?

Pt.: One should talk when you are spoken to.

[This is the closest that the patient comes to a reflective comment about himself and his situation. Yet even at this point he can do no better than phrase it as a proverb, as a bit of collective or folk wisdom rather than as a personal view.]

Dr. C.: Yeah—but well, that is when you are with people who aren't your friends, but you have got to feel friendly towards people. Here we are, just here, we are just talking about these things. You are free to say anything you want to. You will find us pretty understanding people to talk to. —Did this come as a decision on your part not to—I mean, we know that there is nothing wrong with your mouth or your vocal cords, it is possible for you to talk. Now, is this a decision on your part not to talk? —Are you familiar with the old saying: what is it—"you can lead a horse to . . . ." What is the rest of that?

Pt.: ". . . horse to water but you can't make him drink?"

Dr. C.: What do you think of that? Huh? Think there is any truth in it? Huh?

Pt.: Yeah.

Dr. C.: Okay, your patient.

(At this point the interview was terminated. The patient fell asleep about five minutes later, from the effects of the barbiturate.)

CONCLUSION

Like most clinical data, this interview must speak for itself. For some listeners, it may be mute; for everyone, it will point in the direction to which he is inclined. The comments that were made by the writer, as he leaned intrusively over the interviewers' shoulders, are meant to suggest one way of understanding the strange, almost paradoxical phenomenon of a patient whose condition of muteness varied almost from moment to moment. The drug used in this interview is sometimes called "truth serum;" but the selectivity of this patient in regard to the "truth" that he reveals ought to serve as a caution to those who would expect startling results from subjects who are presumably in much better command of their faculties.

It may be, of course, that the critical question asked of the patient—as to why he did not talk—was unanswerable by him for reasons which had little to do with his illness. Perhaps such a question, involving as it does an action as fundamental as language, posed as much difficulty for the patient as would the complementary question, asked of the man in the street, as to why he *does* talk. In any case, the retrospective analysis of case material as attempted here cannot lead us to conclusions, only to some sources of testable questions.

What, then, are the questions which are posed by the clinical data presented here? Three appear to stand out:

1. Why does a powerful sedative make the catatonic patient talk? If his condition is one which is halfway between waking and sleeping, as is claimed by some writers, he should be put firmly to sleep by a sedative and awakened only by a stimulant. What is the nature of a catatonic stupor such that relaxing the patient leads him to begin communicating with others?

2. If, as appears likely, the patient is selective in his silence, we are faced with the issue of will vs. impulse—that is, of whether a specific organization of his impulse life causes him to choose topics he will talk about, or whether he chooses the topics in order to maintain a certain state. A remarkably similar problem has been encountered recently in experimental studies of the phenomenon of subception; whether the presumed nonconscious forces of the personality cause the subject to "recognize" and therefore to avoid certain stimuli before they come to awareness, or whether he selects the stimuli to be avoided and reports his choices in special ways.

3. Now that some information is available, if only inferentially, concerning the nature of the catatonic's experience, how may we use it to frame new generalizations concerning the schizophrenic state?

The proposal offered in this paper, supported by an analysis of the clinical data, is that the schizophrenic steps out of his own history and undertakes to refrain from any commitment to responsible action with others; that he then presents himself, not as a living person with significance to himself and to others, but in the "third person" sense of a set of statistics; and finally, that the catatonic stupor is a state of vigilance in which his declaration of lack of commitment is guarded and preserved. Stated as hypotheses, this proposal as well as all of the other issues raised above, would appear to be amenable to rigorous experimental investigation.

(Received April 23, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

CRITIQUE AND NOTES

A DETERMINER OF THE CENTRALITY OF A TRAIT IN IMPRESSION FORMATION

DAVID T. BENEDETTI AND JOSEPH G. HILL[1]
University of New Mexico

ASCH HAS SHOWN that the impression formed by an experimental subject (*S*), when he is given a list of traits descriptive of a hypothetical person, is not based equally upon the various given traits (Asch, 1946). Some traits are more "central" than others, i.e., they carry more weight in determining the impression than do "peripheral" traits. For example, warm–cold is more central than polite–blunt.

In his studies Asch was not concerned with individual differences among his *S*s. In his words, "We have not dealt in this investigation with the role of individual differences, of which the most obvious would be the effect of the subject's own personal qualities on the nature of his impression" (p. 283).

It is precisely this possibility of relationships among the traits of the perceiver and those of the perceived that prompted the present investigation. In passing, note that the term "perception" is used here in the broad sense of "apperception and cognition," as in Tagiuri and Petrullo (1958, p. x).

A number of other workers have found that the perception of others is related to variables in the perceiver (e.g., Crockett & Meidinger, 1956; Fensterheim & Tresselt, 1953; Lindzey & Rogolsky, 1950; Stagner, 1948). To our knowledge, however, no one has approached this relationship in the context of the Asch method.

This method involves a laboratory procedure that appears remote from the real-life formation of impressions of perceivable persons. This observation is mitigated, however, by an experiment that showed that warm–cold exhibited centrality in a situation which involved an observed, real person (Kelley, 1950).

The present investigators were concerned specifically with the question of whether the observer's possession of a given trait would influence the weight attached to that trait when perceived in another person. Accordingly, the hypothesis was erected that the centrality of a trait varies with the strength of that trait in the perceiver.

METHOD

Subjects. A total of 204 general psychology students at the University of New Mexico were allocated to three groups of 68 as explained below. *S*s were unaware of their classification and of the purpose of the experiment.

Procedure. The trait pair sociable–unsociable was selected for experimental manipulation because a pilot study had indicated that this trait, which is objectively measured in the Gordon Personal Profile[2] (Gordon, 1953), is high in centrality. The Gordon test was administered to all *S*s, who were then divided into High, Middle, and Low on the basis of their Sociability scores. The lowest percentile in the High group was 73, and the highest percentile in the Low group was 37.

Each group was subjected to the same experimental procedure, which followed Asch's method. *S*s were read a series of traits that described a hypothetical person; after the second reading they wrote a short sketch giving their impression of the person. A random half of each group, working under the "Sociable" condition, heard the following traits: intelligent, skillful, industrious, sociable, determined, practical, cautious. The other half of the group worked under the "Unsociable" condition, hearing: intelligent, skillful, industrious, unsociable, determined, practical, cautious. As in Asch's study, the two series were identical except for one varied trait, in this case sociable–unsociable in place of Asch's warm–cold.

The thumbnail sketches of the hypothetical person were procured only for the purpose of maintaining a parallel with Asch's procedure. After the sketches were written the *S*s were given a list, adapted from Asch, of 20 trait pairs, from each of which they selected the trait that best fitted their impression. This session occurred during the nine laboratory periods in the week following the administration of the Gordon test in the two large lecture sections; *S*s were given no hint that the two events were related. The 20 trait pairs are given in Table 1, with the nonpreferred trait in each pair enclosed in parentheses; on the check list given to *S*s, the preferred trait in each pair came first or second at random.

The prediction. Operationally stated, our hypothesis predicted differential checking of preferred traits in the three groups. In terms of the analysis

---

[1] J. G. Hill is now at the Los Alamos Scientific Laboratory.

[2] The following statements are illustrative of the Gordon items measuring sociability: "enjoys having lots of people around"; "a good mixer socially"; "finds it easy to make new acquaintances."

TABLE 1

NUMBERS OF HIGH, MIDDLE, AND LOW SOCIABILITY SUBJECTS ATTRIBUTING PREFERRED TRAITS UNDER THE SOCIABLE AND UNSOCIABLE CONDITIONS

(N of each subgroup = 34)

Trait	High Subjects		Middle Subjects		Low Subjects	
	Soc.	Unsoc.	Soc.	Unsoc.	Soc.	Unsoc.
Generous (ungenerous)	33	5	32	6	32	15
Happy (unhappy)	32	6	31	6	32	22
Good-natured (irritable)	32	5	33	9	31	15
Humorous (humorless)	31	4	30	5	30	3
Popular (unpopular)	33	3	32	4	33	6
Reliable (unreliable)	34	33	34	33	33	34
Humane (ruthless)	31	10	32	17	30	27
Good-looking (unattractive)	32	4	30	10	30	15
Persistent (unstable)	34	30	34	31	34	34
Talkative (restrained)	32	9	33	15	32	20
Altruistic (self-centered)	31	4	19	13	20	14
Imaginative (hard-headed)	29	6	25	13	27	22
Strong (weak)	34	28	32	25	30	26
Honest (dishonest)	33	34	34	34	33	34
Bold (timid)	30	5	32	11	29	21
Mature (immature)	34	10	33	30	32	32
Relaxed (tense)	32	6	27	9	24	10
Decisive (indecisive)	34	29	32	31	31	34
Broad-minded (narrow-minded)	33	5	31	19	31	21
Secure (insecure)	33	10	29	19	31	27
Mean	32.35	12.30	30.75	17.00	30.25	21.60

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE

Source	df	Mean Square	F
Groups	2	130.44	
Conditions	1	6006.67	18.42
Groups × Conditions	2	326.10	6.35*
Within Groups	114	51.34	

* $p < .01$.

of variance, a significant F either for Groups or for Groups × Conditions interaction would confirm this prediction.

RESULTS

Table 1 gives the number of *S*s in each subgroup who checked the preferred traits, and Table 2 gives a summary of the analysis of variance of these data.

The significant F for the Groups × Conditions interaction confirms the prediction: the classification of the *S*s by sociability was a source of real differences which were contingent upon the condition under which *S*s worked. Note, in Table 1, that the mean number of *S*s attributing the preferred traits under the Sociable condition remained virtually constant across the three groups (32.35, 30.75, 30.25), while under the Unsociable condition the corresponding means rose from 12.30 in the High group, through 17.00 in the Middle group, to 21.60 in the Low group.

Neither Groups nor Conditions achieved significance independently. The F for Conditions, as tested against Interaction since the latter is significant (see Dixon & Massey, 1951, pp. 137–138, or Walker & Lev, 1953, p. 360), falls barely short of the value required for significance at the .05 level; in any event, Conditions as a primary variable was of no concern.

DISCUSSION

The hypothesis that the centrality of a trait varies with the strength of that trait in the perceiver receives qualified support with respect to the trait pair sociable–unsociable as tested in the present manner.

The measured sociability of our *S*s was significantly related to the impressions they formed, defined in terms of the frequency with which preferred traits were attributed to the described person. This relationship was, however, confined to the situation in which the person was described as unsociable. Under this condition, *S*s formed impressions that varied in the High, Middle, and Low Sociability groups: while Low Sociability *S*s were relatively generous in assigning preferred traits to the "unsociable" person, the High Sociability *S*s were more reluctant to do so, and the Middle Sociability *S*s intermediate.

The description of the hypothetical person as sociable produced in all three groups an impression so favorable that the sociability classification of the *S*s was, so to speak, washed out.

Whether this asymmetrical effect obtains in connection with traits other than sociable–unsociable remains for future research to determine.

The results of the present study, in indicating that one's own sociability may influence the centrality of that trait in the forming of impressions of others, enhance the possibility that "person perception" may involve other relationships between the traits of the perceiver and those of the perceived regarding the saliency and weight of those traits in the resulting impressions.

REFERENCES

ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.*, 1956, **53**, 378–380.

DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.

FENSTERHEIM, H., & TRESSELT, H. The influence of value systems on the perception of people. *J. abnorm. soc. Psychol.*, 1953, **48**, 93–98.

GORDON, L. V. *Gordon personal profile*. New York: World Book, 1953.

KELLEY, H. H. The warm–cold variable in first impressions of persons. *J. Pers.*, 1950, **18**, 431–439.

LINDZEY, G., & ROGOLSKY, S. J. Prejudice and identification of minority group membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 37–53.

STAGNER, R. Psychological aspects of industrial conflict: I. Perception. *Personnel Psychol.*, 1948, **1**, 131–144.

TAGIURI, R., & PETRULLO, L. (Eds.) *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univer. Press, 1958.

WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received January 26, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

THE PSYCHOLOGICAL MEANING OF ACQUIESCENCE SET:

FURTHER EVIDENCE

N. L. GAGE AND B. B. CHATTERJEE

Bureau of Educational Research, University of Illinois

THIS PAPER strengthens an earlier hypothesis (Gage, Leavitt, & Stone, 1957) by means of a more appropriate method and correspondingly more significant results. The hypothesis was that acquiescence set—the tendency to say Yes or Agree when in doubt—partakes of "authoritarian submission" or "a submissive uncritical attitude toward idealized moral authorities of the ingroup" (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, pp. 230–232).

From this hypothesis it was reasoned that either Agree or Disagree responses to positive items—items expressing nonauthoritarian beliefs or attitudes—would be equivocal in their significance. "Logically," agreement with such items would betoken nonauthoritarianism. "Psychologically," however, agreement with such items, especially when the statements were vague and ambiguous, might reflect the authoritarian submission trend. The same ambiguity would characterize Disagree responses to positive items.

On the other hand, an Agree or Disagree response to a negative item—a statement expressing authoritarian beliefs or attitudes—would be unequivocal in its significance for the personality of the respondent. Both "logically" and "psychologically," by making the basic assumption, Agree responses could be considered to indicate authoritarianism, while Disagree responses would unequivocally signify the opposite.

In short, responses to positive items could mean one thing logically and the opposite psychologically, while responses to negative items would mean the same thing both logically and psychologically. Accordingly, it was predicted that *negative items have greater validity than positive items for the measurement of authoritarianism.*

In the paper referred to, this hypothesis was tested not with the F scale, which contains only negative items, but with a similar instrument, the Minnesota Teacher Attitude Inventory (MTAI) (Cook, Leeds, & Callis, 1951). The MTAI contains 38 positive items and 112 negative items. As an external, highly reliable, and readily available but far from perfect criterion of authoritarianism against which to determine the validity of the MTAI items, the original study used pupils' ratings of their teachers on a 48-item "My Teacher" inventory (Leeds & Cook, 1947, p. 154).

In this original study, the hypothesis was supported, but only weakly. The validity of the positive items, for 97 teachers of Grades 4–6, was .13; the three randomly made sets of negative items had a mean validity of .25. The difference between the two r's of .13 and .25 was reported as barely significant at the .05 level, using Hotelling's t test (1940). In a subsequent computation of the significance test, the mean (.69) of the three separate r_{23}'s between the positive set and the three sets of negative items was used, instead of the single r_{23} (.73) between the positive set and the total set of 112 negative items. This change in r_{23} made the difference between r_{12} and r_{13} slightly less significant; i.e., as is shown for the Della Piana-Gage

TABLE 1

STATISTICAL SIGNIFICANCE OF THE REPLICATIONS AND COMBINED RESULTS

Source of Data	r_{12} (Validity, Pos. Items)	$\bar{r}_{13}$ (Mean, 3 Neg. Item Validities)	$\bar{r}_{23}$ (Mean r, Pos. Items vs. 3 Neg. Item Sets)	$t_{(r_{13}-r_{12})}$	df_t	p (one-tail)	χ^2 Equiv. of p*	df_{χ^2}
Della Piana-Gage	.13	.25	.69	1.53	94	.07	5.3185	2
Mazzitelli	.18	.12	.62	−.78	123	.75	.5754	2
Cook-Hoyt	.14	.25	.62	1.04	64	.14	3.9322	2
Leeds	.24	.27	.51	.31	97	.37	1.9885	2
Combined results							11.8146*	8

*p = .17.

data in Table 1, t was now 1.53 instead of 1.6,[4] and the one-tailed p was now .07 instead of .05.

THREE REPLICATIONS

This method of testing the hypothesis had a major weakness: it employed positive and negative items that had been highly selected on the basis of empirical evidence of their validity in discriminating between teachers on the criterion measure. Such empirical selection of items should tend to equalize the validity of the positive and negative items and make any difference in their validity hard to demonstrate. This limitation of method showed itself in three replications of the original study.

Highly comparable sets of data were available from studies by Mazzitelli (1957), Cook and Hoyt[1] (1958), and Leeds (1946); these authors generously made their data available to us. The results of these replications, together with that of the original study, are shown in Table 1. As before, the number of positive items was 38, and the 112 negative items were divided at random into three sets of 38, 38, and 36. As is shown in Table 1, three of the four studies yielded results in the direction of the hypothesis, but none of the obtained differences between r_{12} and r_{13} reached the .05 significance level. When the chi square model for testing the significance of combined results was applied, p was .17 (Jones & Fiske, 1953).

Clearly, it seems unlikely that a statistically significant difference in validity will be found between the positive and negative items of the published version of the MTAI. Our explanation of this is that the MTAI's positive and negative items have been highly selected and hence to some extent equated, on the basis of their validity in differentiating among teachers on a variable substantially like authoritarianism.

A MORE APPROPRIATE METHOD

Rather than reject the hypothesis on the basis of these results, we sought data more appropriate for a test of it. Such data proved to be available in Leeds' (1946) doctoral dissertation. In the course of that study, Leeds wrote 754 items aimed at eliciting attitudes of teachers relevant to their ability to get along with their pupils on nonauthoritarian terms. He divided these 754 items into two forms of a try-out instrument, Forms A and B. Each form was composed of approximately equal numbers of positive and negative items. Furthermore, most of the positive items on Form A reappeared in negative versions on Form B, and most of the negative items of Form A reappeared in positive versions on Form B.

These two forms were administered by Leeds some weeks apart to 200 teachers of Grades 4–6. One hundred of these teachers had been identified by their principals as highly superior in getting along with pupils, in maintaining a good classroom atmosphere, and so on; the other 100 had been identified by their principals as extremely inferior in these same respects.

Leeds made available distributions of the superior and inferior teachers' responses to 738 of the 754 items. It was hence possible to compute a validity index for these items (Mazzitelli, 1957). This was done by assigning weights of 1–5 to each of the five response alternatives (agree strongly, agree, undecided, etc.) and obtaining two sums of these weights, one for the responses to a given item by the 100 superior teachers and the second for the responses by the 100 inferior teachers. The index of item validity was the absolute difference between the sum of the response weights for the 100 superior teachers and that for 100 inferior teachers.

Prior to this computation of validity indices, each item was classified as "positive" or "negative," i.e., as stating a nonauthoritarian or an authoritarian belief or attitude, by two graduate students in education who had been given brief

---

[1] W. W. Cook & C. J. Hoyt, Personal communication, 1958.

TABLE 2

VALIDITY INDICES OF POSITIVE AND NEGATIVE ITEMS IN LEEDS' FORMS A AND B

Validity Index	Form A		Form B	
	Positive Items	Negative Items	Positive Items	Negative Items
75–79		1	1	
70–74				
65–69		1		1
60–64				
55–59	1	1		
50–54	3	4		1
45–49	4	13	2	7
40–44	3	6	2	8
35–39	6	9	2	17
30–34	5	22	8	15
25–29	18	22	18	28
20–24	22	29	21	21
15–19	30	22	16	20
10–14	26	23	34	34
5–9	19	21	22	30
0–4	39	17	39	24
No. of items	176	191	165	206

TABLE 3

CHI SQUARE OF JOINT DISTRIBUTIONS OF ITEM VALIDITY INDICES AND POSITIVE–NEGATIVE CHARACTER OF ITEM

	Form A		Form B	
	Pos. Items	Neg. Items	Pos. Items	Neg. Items
Higher Validity	62	108	70	118
Lower Validity	114	83	95	88
	$\chi^2 = 16.75$,	$p < .001$;	$\chi^2 = 8.12$,	$p < .01$

definitions of these categories. These two judges agreed in their classification of 707 or 94 per cent of the 754 items. The 47 items on which the judges disagreed were categorized by a third judge (Gage), without knowledge of their validity indices.

Table 2 shows the frequency distributions of these validity indices for the positive and negative items of Forms A and B. Note that for both forms the central tendency of the negative items is obviously higher than that of the positive items. Splitting the two distributions of validity indices for each form at approximately the median of the combined distribution for that form yields the 2 × 2 tables shown in Table 3. Chi square is 16.75 (p less than .001) for Form A and 8.12 (p less than .01) for Form B.

SUMMARY AND CONCLUSIONS

This paper provides further evidence on the hypothesis that "negative" items—those stating authoritarian beliefs or attitudes—are more valid than "positive" items for measuring authoritarianism. First, it is shown that it is inappropriate to test this hypothesis with the highly selected positive and negative items of the published version of the Minnesota Teacher Attitude Inventory. A test of the combined significance of four sets of results using the published MTAI failed to yield a probability low enough to cast doubt upon the null hypothesis.

A more appropriate test of the hypothesis would compare the validity of positive and negative items not previously selected on the basis of their validity. Such a comparison was carried out using 738 items originally written by Leeds in the process of constructing the MTAI. The criterion of item validity consisted of principals' judgments of the effectiveness of teachers in maintaining a nonauthoritarian classroom atmosphere. The results for both Forms A and B, each consisting of positive and negative items in approximately equal proportions, are consistent with the hypothesis. In both forms, the proportion of the negative items with validities above average is very significantly higher than that of the positive items.

Further replication, employing the best available sets of positive and negative F scale items (e.g., Christie, Havel, & Seidenberg, 1958), is in order. The validity of the items should of course be estimated against one or more *external* criteria of authoritarianism. By this method, rather than by mere internal analyses of relationships between positive and negative items themselves, more conclusive evidence can be obtained on the issue whether acquiescence can be identified *in part* with authoritarianism. Such replication will shed light on whether the superior validity of negative items here found stems from the item-writing idiosyncrasies of one test author (Leeds) or whether it reflects a basic similarity in psychological meaning between acquiescence set and the authoritarian submission trend.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

COOK, W. W., LEEDS, C. H., & CALLIS, R. *The Minnesota Teacher Attitude Inventory: Manual.* New York: Psychological Corp., 1951.

GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescence set for authoritarianism. *J. abnorm. soc. Psychol.*, 1957, **55**, 98–103.

HOTELLING, H. The selection of variates for use in prediction with some comments on the problem of nuisance parameters. *Ann. math. Statist.*, 1940, **11**, 271–283.

JONES, L. V., & FISKE, D. W. Models for testing the significance of combined results. *Psychol. Bull.*, 1953, **50**, 375–382.

LEEDS, C. H. The construction and differential value of a scale for determining teacher-pupil attitudes. Unpublished doctoral dissertation, Univer. of Minnesota, 1946.

LEEDS, C. H., & COOK, W. W. The construction and differential value of a scale for determining teacher-pupil attitudes. *J. exp. Educ.*, 1947, **16**, 149–159.

MAZZITELLI, D. A forced-choice approach to the measurement of teacher attitudes. Unpublished doctoral thesis, Univer. of Illinois, 1957.

(Received January 5, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

HEAD–BODY DIFFERENTIATIONS IN BODY IMAGE AND SKIN RESISTANCE LEVEL[1]

SEYMOUR FISHER

Baylor University College of Medicine and Houston State Psychiatric Institute

PREVIOUS STUDIES have shown that the definiteness an individual ascribes to his body boundaries is correlated with the relative physiological reactivity of his body exterior vs. his body interior (Fisher & Cleveland, 1958). Similarly, it has been demonstrated that the size or strength attributed by the individual to his right vs. left body sides is correlated with the relative GSR reactivity of the two sides (Fisher, 1958, 1959). Such findings have been interpreted as supporting the theory that body image attitudes influence the reactivity of body areas.

The present study was undertaken to investigate a new dimension of body image and its relationship to the reactivity of given body areas. This new body image dimension involves the character of the distinction between the head area and the non-head area. It was conjectured on the basis of previous observations (Fisher & Cleveland, 1958; Reich, 1949) that one of the basic problems in body image organization relates to the degree of importance to be assigned to the head vs. the body in the over-all body schema. It was therefore hypothesized that there are individual differences such that at one extreme the head is experienced as relatively large and the body area as relatively small and at the other extreme the opposite pattern prevails. Further, it was hypothesized that the greater the relative size attributed to one of these sites as contrasted to the other, the greater is its physiological reactivity with reference to the other.

METHOD

Body Image Measure

The measure of head–body differentiation was based on size estimates involving head width and body width. Each *S* was taken into a dark light-proof room and asked to stand 6 ft. from two vertical luminous rods which were on a base 40 in. from the floor. The rods were 23 in. in length and ⅛ of an in. in width. They were initially set so as to be touching, without an intervening space between them. Each *S* was told that there were two rods and that they would be gradually moved apart. *S* was to indicate when he judged the space between them to be equal to the difference between the width of his head and the width of his body. Obviously, the greater the space between the rods the smaller was the magnitude he assigned to the head in the head–body comparison. The index used was the average of three judgment trials.

The precedent for using size judgments to determine the relative prominence of body areas in the body schema has already been established in previous studies (Fisher, 1958, 1959; Fisher & Abercrombie, 1958). It should be emphasized that such size judgments have been successful only when obtained in a setting which maximizes the projection of the autistic components of body size concepts. Projection was encouraged in the present study by obtaining judgments in a dark room where *S*s could not use ordinary visual cues from their own bodies or those of other individuals as an anchoring frame of reference. It was encouraged also by the fact that instead of requesting *S*s simply to estimate the sizes of their heads or

---

[1] Supported by USPH Career Research Investigator Grant (M 1578).

bodies separately, they were asked to make a more complex decision which involved not only an estimate of the magnitude of each area, but also a simultaneous comparison of them. Preliminary observations had shown that *S*s made much more unrealistic judgments under these comparison conditions than under conditions calling for separate estimates of head and body width.

Head Area–Body Area Reactivity

Skin resistance was used as the measure of relative physiological reactivity of head region vs. non-head region. A Brush direct writing oscillograph was employed. A constant current supply of 20 microamperes and a DC amplifier were used for measuring the voltage across the *S*. The record was calibrated directly in ohms. Separate balanced systems were utilized for the head and non-head measures. The area of recording from given sites was made equivalent by using pieces of tape with two holes, each of ¼ in. diameter, punched into them. Two minutes of recording were secured simultaneously from any pair of head vs. non-head sites.

The non-head area recording was taken from the dorsal surface of the middle segment of the middle finger of the left hand. This arbitrary choice was based on such factors as ease of applying electrodes and similarity of the skin in this area to facial skin. More central portions of the body were not evaluated because of the practical difficulties of dealing with regions which are covered by clothing in both sexes.

Three separate recording points were designated for the head. Site 1 was ½ in. directly above the glabella. Site 2 was on the left cheek, 2½ in. directly lateral to the nasal septum. Site 3 was in the middle of the chin, half way between its lowest point and the lower lip of the mouth. Three sites were used for the head because there were individual variations in facial contours which made site determination unreliable. The resistance readings were taken in turn from each of the head locations simultaneously with readings from the finger.

An index descriptive of relative head–body reactivity was formulated that was equal to the ratio of body resistance to head resistance at that point in the recording where the maximum difference between the two occurred. This index was developed on the basis of preliminary exploration which indicated that it correlated better with the head–body differentiation measure than did other resistance level representations.

The test-retest reliability of the resistance index was evaluated by retesting 15 *S*s. The intervals between test and retest ranged from one day to three months, with a median of two months. A rho of .87 between forehead–hand reactivity ratios was obtained.

Subjects

There were 62 *S*s, 18 female and 44 male. The median age was 21, and the median educational level was 15 years. Most of the *S*s were recruited from sororities and fraternities and paid for their services. A small portion of the group included secretaries and laboratory technicians.

RESULTS

Scores for both the reactivity measures and the body image measures varied widely. The finger-head reactivity ratios ranged from 40-to-1 to 1-to-7. The median ratio was one in which the finger resistance was three times as great as the head resistance.

*S*s' estimates of the difference in width between head and body varied from a low of 2 in. to a high of 17. The median was 6 in.

Chi square indicated significant relationships in the predicted direction between the body image measure of head–body differentiation and the skin resistance ratio indices. As shown in Table 1, *S*s above the median in degree of relative size attributed to the head manifested significantly smaller ratios of head area resistance to non-head area resistance than did *S*s below the median. This was true at the .001 level for the forehead–hand sites, between the .02 and .01 levels for the cheek–hand sites, and at the .001 level for the chin–hand sites. The clearest relationship between the body image measure and the reactivity ratio emerged when the reactivity ratio was computed on the basis of the forehead–hand sites. Perhaps this was due to the fact that electrode placement is easier to standardize on the forehead than on the cheek or chin.

It was noted that a larger proportion of the variance of the reactivity ratios contributing to the significant relation with the body image

TABLE 1
CHI SQUARE ANALYSIS OF RELATIONSHIP OF BODY IMAGE MEASURE TO VARIOUS HEAD–BODY REACTIVITY RATIOS

Variables	Chi Square	Significance Level
Body Image vs.		
Forehead–Hand Ratio	9.2[a]	.001
Cheek–Hand Ratio	5.8	.02–.01
Chin–Hand Ratio	8.4	.001

[a] The direction of the results in each instance is such that the greater the size attributed to the head relative to the body, the lower is the resistance level of the head site as compared to the non-head site.

measure was derived from the non-head skin resistance than from the head resistance. Thus, chi square indicated only a chance relationship between forehead skin resistance alone and the body image index. However, the skin resistance of the hand alone was linked at the .001 level with the body image measure; thus, the greater the relative size ascribed to the head, the greater was the hand skin resistance. The question therefore arose of whether the non-head resistance measure alone might be the equivalent of the ratio index. A rank order correlation between the ratio index (forehead–hand) and the non-head resistance measure alone yielded a value of .69. Further, the ratio index (forehead–hand) was correlated (rho) .29 (.01 level) with the body image measure; the analogous correlation for the non-head resistance measure was .34 (.01 level). It consequently appears that the ratio index and the non-head resistance measure are similar. They are not equivalent, however, and the ratio index seems to be slightly better correlated with the body image variable than the non-head resistance measure alone. For these reasons and because it makes more sense in theoretical terms, the ratio index was considered the most appropriate representation of physiological reactivity.

DISCUSSION

The data support the hypothesis that different levels of reactivity in given body sectors are linked with corresponding differences in body image attitudes toward such sectors. Thus, when the head area was perceived as having large magnitude with reference to the body, it was characterized by relatively lower skin resistance (indicating greater reactivity) than that of the non-head area. But when the head area was judged as being small with reference to the body, it was found to have relatively higher skin resistance. These findings are congruent with previously reported correlations between body image boundary definiteness and body interior vs. body exterior reactivity and also between body image measures of right and left side body differentiation and right vs. left GSR reactivity. It is a speculative step to deduce from the significant relationships between body image and body reactivity that body image attitudes toward a given body sector may play a causative role in the level of reactivity of that sector. The significant relationships could possibly be due to some third variable, or varying levels of reactivity in body areas may give rise to differential judgments of comparative size.

The possibility that some third variable plays a major part in the observed relationships must continue to be seriously considered. However, the idea that body image differentiations between body sectors simply grow out of sensation or reactivity differences between these sectors is strongly questioned by two previous findings. First, it has been shown (Fisher & Cleveland, 1958; Fisher & Fisher, in press) that the individual's attitudes toward his body sectors (viz., inside vs. outside and right vs. left) are significantly linked with past socialization experiences as they are reflected in his images of the parental figures. It is difficult to see why such a relationship could exist if a body image attitude toward a localized body area were due simply to the reception of body sensations from that area. Second, it has been demonstrated (Fisher & Cleveland, 1958) that the mere experiencing of intense sensations from a body area (e.g., persistent muscle pain), even for a period of years, does not significantly affect basic body image attitudes of the sort which have here been under consideration.

It is becoming increasingly clear that there are not one or two important dimensions of body image but rather a complex array. There is also a corresponding array of differences in physiological reactivity between various body sectors. The task remains of determining the full range of important body image dimensions and how they are related to each other.

SUMMARY

This study concerned itself with testing the hypothesis that the relative physiological reactivity of the head area vs. the non-head area is linked with the relative size representation assigned to each in the body schema. Size estimates concerning the head and body were obtained from 62 *Ss* in an unstructured situation. Also, skin resistance levels were obtained from head and non-head sites simultaneously. The over-all results were clearly supportive of the hypotheses.

REFERENCES

FISHER, S. Body image and asymmetry of body reactivity. *J. abnorm. soc. Psychol.*, 1958, **57**, 292–298.

FISHER, S. Body reactivity gradients and figure drawing variables. *J. consult. Psychol.*, 1959, **23**, 54–59.

FISHER, S., & ABERCROMBIE, J. The relationship of body image distortions to body reactivity gradients. *J. Pers.*, 1958, **36**, 320–329.

FISHER, S., & CLEVELAND, S. E. *Body image and personality.* Princeton, N. J.: Van Nostrand, 1958.

FISHER, S., & FISHER, RHODA. A developmental analysis of some body image and body reactivity dimensions. *Child Develpm.*, in press.

REICH, W. *Character analysis.* New York: Orgone Inst. Press, 1949.

(Received November 26, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

EXTRINSIC RELIGIOUS VALUES AND PREJUDICE

W. CODY WILSON

Harvard University

A GROWING body of empirical evidence indicates that people who are "religious" according to various criteria tend to be more prejudiced than those who are not "religious" (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950; Allport & Kramer, 1946; Levinson, 1954; Pettigrew, 1956; Rosenblith, 1949; Stouffer, 1955; Turbeville & Hyde, 1946).

Adorno, et al. (1950), as a result of qualitative analysis of some of their data, suggested that the fact of acceptance or rejection of religion is not as important as the way in which it is accepted or rejected (p. 217); they distinguished between a religion that is conventional, externalized, and neutralized and one that is more personal and internalized. Allport (1954) also emphasized the motivation for religious affiliation, postulating that those who affiliate with a religious institution for utilitarian, self-serving purposes—those with "extrinsic" religious values—tend to be prejudiced.

Levinson, following the leads developed in *The Authoritarian Personality*, devised a Religious Conventionalism Scale which records the extent of agreement or disagreement with statements of conventional religious beliefs about the church, the Bible, God, prayer, etc. Correlations between this scale and the California Ethnocentrism Scale of the general magnitude of .50 were reported (Levinson, 1954) for samples heterogeneous in terms of religious affiliation.

This paper reports data concerning another scale of religious values, designed for use with individuals affiliated with a religious institution and reflecting the ideas of Adorno, et al. (1950) and Allport (1954), and compares it with the Religious Conventionalism Scale.

The Extrinsic Religious Values Scale[1]

The instrument is composed of 15 dichotomous items.[2] These items are of two general types: (*a*) one subgroup reflects an allegiance to, and dependence upon, the external or institutional structure of a church; and (*b*) the other subgroup reflects a utilitarian orientation toward religion, i.e., acceptance of religion as a means. One of the two alternative responses to each item reflects an "extrinsic religious value"; the other alternative is not positively defined but simply reflects the absence of extrinsic religious value. The score on the instrument is the sum of the items for which the "extrinsic" alternative is preferred.

Two examples of the items of the Extrinsic Religious Values scale (the asterisk indicates the alternative that reflects the extrinsic value) are: Prayer is, above all else, a means of obtaining needed benefits, protection, and safety in a dangerous world.

_*_ On the whole, I agree.
___ On the whole, I disagree.

Some people say that they can be genuinely religious without being a member of any church.

___ Yes, I agree that they can.
_*_ I do not believe this.

The scale has a corrected equivalent-halves reliability of between .80 and .85 for homogeneous Protestant samples with *N*s in the range from 15 to 35.

METHOD

The Extrinsic Religious Values Scale (ERV), a 12-item version of Levinson's Religious Conventionalism Scale (RC), a 12-item version of the California Anti-Semitism Scale (AS), and filler items were assembled into a questionnaire with typical instructions. The total questionnaire contained approximately 50 items and required about 25 min. for 90% of people with a high school education to complete. The questionnaire was administered to the following small, relatively homogeneous groups at regularly scheduled organizational meetings:

1., 2., 3., 4. Adult members of four different Congregational Christian churches in the metropolitan Boston area (N = 15, 22, 22, and 35 respectively);
5. Members of a young adult organization in a Methodist church in the Boston area ($N = 10$);
6. Harvard-Radcliffe students of various Christian denominations who reported attending services regularly ($N = 26$);
7. Adult members of a Disciples of Christ church in North Carolina ($N = 11$);
8. Adult members of the choir of a Negro Baptist church in the Boston area ($N = 15$);
9. Sophomore students in a Catholic girls college in the Boston area ($N = 23$);
10. Catholic graduate students in the Boston area ($N = 28$).

---

[1] The aid of Kay Wilson and G. W. Allport was invaluable in the formulation of these items.

[2] The Extrinsic Religious Values Scale has been deposited with the American Documentation Institute. Order Document No. 6135, remitting $1.25 for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies.

TABLE 1

RELATIONSHIP BETWEEN EXTRINSIC RELIGIOUS VALUES, RELIGIOUS CONVENTIONALISM, AND ANTI-SEMITISM

Group	N	Scale reliability[a]	ERV		RC		AS		Correlations[c]		
			Mean	*SD*	Mean	*SD*	Mean	*SD*	ERV-AS	RC-AS	ERV-RC
1	15	.80	7.0	3.0	68.1	9.6	48.0	21.8	.65**	.39	.51*
2	22	.82	6.3	2.0	—	—[b]	36.0	12.5	.66***	—[b]	—[b]
3	22	.72	7.3	2.0	64.5	10.1	44.3	11.9	.56**	.08	.08
4	35	.85	7.5	2.6	60.5	9.5	47.7	15.9	.68***	.29	.20
5	10	.81	5.2	1.8	—	—[b]	35.0	11.2	.68*	—[b]	—[b]
6	26	.85	5.0	1.9	52.5	17.7	33.2	12.7	.71***	.41*	.17
7	11	.86	7.9	2.3	—	—[b]	51.6	15.1	.72*	—[b]	—[b]
8	15	.82	9.2	2.5	69.0	8.4	41.0	10.8	.64**	.31	.23
9	23	.57	6.6	1.7	78.2	6.3	40.8	10.7	.41*	.30	.40*
10	28	.51	6.1	1.7	75.9	5.5	36.0	9.1	.43*	.42*	.57**

Note.—The following abbreviations are used in the column headings: ERV, Extrinsic Religious Values Scale; RC, Religious Conventionalism Scale; and AS, Anti-Semitism Scale.

[a] Corrected equivalent halves.

[b] The Religious Conventionalism Scale was not administered to this sample.

[c] All correlations are by rank-difference method.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

It should be noted that these groups, totaling 207 subjects, cover a wide range of religious, ethnic, and geographic variation.

RESULTS

The relation between the Extrinsic Religious Values Scale and the Anti-Semitism Scale was positive and significantly greater than zero for each of the 10 samples as shown in Table 1.

The Extrinsic Religious Values Scale has a consistently larger relationship to the Anti-Semitism Scale than does the Religious Conventionalism Scale (columns 10 and 11, Table 1); the ERV-AS correlation is greater than the RC-AS correlation for each of the seven groups that responded to all three scales ($P = .016$, sign test).

DISCUSSION

The reliability of the ERV scale, and consequently its relation to AS, is lower for the Catholic groups than for the Protestant samples (see columns 3 and 10, Table 1). This result is attributable to the fact that approximately one-fourth of the items in the ERV scale did not differentiate among the Catholics; official church doctrine dictates the responses to these items for individuals affiliated with it. The ERV dimension apparently may be measured, and it does correlate with AS, within the Catholic groups, but the scale needs to be revised for use with this population.

The ERV-AS correlation, for the eight Protestant groups, ranged from .56 to .72 with a mean of .66. It may be expected that the actual or "true" relationship between extrinsic religious values and prejudice is considerably larger than these figures. If corrections for homogeneity and attenuation based on mean values in these areas are applied to the mean correlation between ERV and AS, the corrected coefficient is .80. Further, a qualitative inspection of the data indicated that the individuals who contributed the most to the lowering of the relationship between ERV and AS were several who scored relatively high on the ERV scale but extremely low on the more direct AS scale. The reverse, that is, low ERV and high AS scores, did not occur nearly so frequently. It is likely that the validity of the ERV scale is greater than that of the AS scale because its content is more ambiguous in terms of social acceptability. The relationship between extrinsic religious values and prejudice is, therefore, perhaps greater than indicated by the relationship between the two scales.

The Extrinsic Religious Values Scale appears to be twice as effective as the Religious Conventionalism Scale in terms of predicting, or accounting for variance in, prejudice within Protestant groups. If the correlations between ERV and AS, and RC and AS (Columns 10 and 11, Table 1), are transformed into $\sqrt{1 - r^2}$ or r^2, then the value for ERV-AS is more than twice that for RC-AS in all five of the Protestant groups that responded to all three of the scales ($P = .031$, one-tailed sign test).

It may be noted that the RC-AS correlations reported in this study are smaller than the correlations between RC and prejudice previously reported in the literature. This study used samples rather homogeneous with respect to religion and a specific measure of prejudice, where previous studies typically used samples heterogeneous with respect to religion and a generalized prejudice scale.

Perhaps the most surprising finding of this study, in view of the fact that both the ERV and RC scales were based in part on the insights developed in *The Authoritarian Personality*, is the low relationship between the ERV and the RC scales (column 12, Table 1). On the other hand, as the name implies, the Religious Conventionalism Scale concentrates upon the content of religious beliefs whereas the Extrinsic Religious Values Scale attempts to measure more directly the motivations for affiliating with a religious institution.

REFERENCES

ADORNO, T. W., FRENKLE-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ALLPORT, G. W. *The nature of prejudice*. Cambridge: Addison Wesley, 1954.

ALLPORT, G. W., & KRAMER, B. W. Some roots of prejudice. *J. Psychol.*, 1946, **22,** 9–39.

LEVINSON, D. J. The intergroup relations workshop: its psychological aims and effects. *J. Psychol.*, 1954, **38,** 103–126.

PETTIGREW, T. F. Regional differences in anti-Negro prejudice. *J. abnorm. soc. Psychol.*, 1959, **59,** 28–36.

ROSENBLITH, JUDY F. A replication of "Some roots of prejudice." *J. abnorm. soc. Psychol.*, 1949, **44,** 470–489.

STOUFFER, S. A. *Communism, conformity, and civil liberties*. Garden City, New York: Doubleday, 1955.

TURBEVILLE, G., & HYDE, R. E. A selected sample of attitudes of Louisiana State University students toward the Negro: a study of public opinion. *Social Forces*, 1946, **24,** 447–450.

(Received March 16, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 2, 1960

ASSIMILATION OF THE KNOWLEDGEABLE NEWCOMER UNDER CONDITIONS OF GROUP SUCCESS AND FAILURE[1]

ROBERT C. ZILLER AND RICHARD D. BEHRINGER[2]
University of Delaware

WHETHER members of a group perceive a newcomer as facilitating or obstructing may be determined, in part at least, by the needs of the group (Bruner & Postman, 1949). It may be assumed that members of successful groups in contrast to those of groups that have experienced failure perceive their group as a self-sufficient, self-contained, and complete social unit. However, the introduction of a new member may precipitate examination of and changes in the organizational structure, the methods of sharing group rewards, and the group's modes of adjustment to its environment. Thus, to a successful group, the newcomer represents a potentially disruptive force threatening to stimulate changes in positively reinforced behavior. Unsuccessful groups, on the other hand, may be expected to perceive the newcomer as a possible resource and as an agent or catalyst of change. Thus, it was hypothesized that newcomers are accepted less readily by successful groups than by groups with a history of failure.

METHOD

Subjects

The subjects (*S*s) were male and female volunteers from freshman English classes at the University of Delaware. The 120 *S*s were assembled at random in 40 three-person groups homogeneous in sex.

Procedure

The experiment was described to the *S*s as a study of group decision-making. The three-person host groups were required initially to solve two problems, a Meier Art Judgment problem and a dot-estimate problem, through the medium of a standard note-passing communication net of the sort used by Bavelas. Following an induction of group success or failure a new member was introduced to the group on the pretext of studying the effects of group size on decision-making. At this point, the new group was presented with a second dot-estimate task that was very much more difficult than the first. The newcomer was, in fact, a confederate who knew the correct answer and a correct method for arriving at the answer. The criteria according to which assimilation of the newcomer was evaluated included the degree to which the group was persuaded to accept the correct answer, the degree to which the members felt that they had changed their initial estimates, three ratings of the newcomer by the host group members, and the number of notes addressed to the newcomer.

---

[1] This report is an extension of a paper read before the American Psychological Association meetings in Cincinnati, Ohio, 1959.

[2] Most of the statistical analyses were conducted by Paula Tammany.

The inductions of success or failure followed the two initial tasks. Failure-groups were informed that their decision on the Art Judgment task was wrong and that on the dot-estimate task the group performed "only better than 10% of the groups who have participated in this task in the past . . . that is, nine out of ten groups that we have previously tested performed better than your group. . . ." Similarly, a "successful" group was informed that their decision on the art task was correct and that the group performed better than 90% of the groups on the dot task.

The dot display used for the third and critical problem presented ten times as many dots as in the first task, contained in an irregular figure resembling a ping-pong paddle rather than the rectangular figure previously employed. The slide was exposed for only five sec. First, the members submitted individual judgments as to the correct number of dots. Next, the group was permitted "fifteen minutes in which to reach a group decision as to the correct number of dots on the card." Finally, the members each submitted private estimates of the "number of dots they personally really thought there were on the slide."

The confederates (the newcomers) received detailed instructions about their roles. Their initial message to all members laconically communicated the correct "estimate" and a correct arithmetic process for arriving at it. Thereafter, the confederates merely responded to inquiries with prescribed messages reflecting profound confidence in their estimate but willingness to compromise for the sake of group efficiency and harmony. Moreover, they were enjoined not to submit the final group estimate. It should be emphasized that the task was of such complexity that the naive Ss could neither estimate the correct number of dots with confidence nor discover the correct approach.

The method used by the confederates required them to focus on the rectangle of minimum dimensions which enclosed the paddle-shaped dot display. The paddle-shaped figure covered approximately one-half the area within the rectangle. Thus, the number of dots within the figure could be approximated by estimating the width and length of the rectangle in dots and dividing the product by two.

Measures

The dependent variables presumed to reflect degree of assimilation included measures of communication direction, group influence, and the relative ratings of the newcomer by the host-group members (Ziller, Behringer, & Jansen, 1959).

The measure of communication direction was simply the total number of messages sent by the naive *S*s to the newcomer.

The first measure of group influence was the absolute difference between the group decision and the correct answer. The second measure of group influence was derived from the ratio of the degree of change in estimate in the direction of the correct answer to the degree of possible change. An upper limit of 1.00 is inherent in the index and a lower limit of zero was imposed. The resulting ratio was subjected to the arc sine transformation. The mean of the members' transformed indices provided the index of estimate change for a group.

A phenomenological measure of influence was derived from the mean of the weighted responses of the naive members to the item: "I changed my individual estimate of the number of dots a great deal as a result of the group discussion." The alternatives were arranged on a six-point scale varying from "agree very much" to "disagree very much."

Finally, the naive members rated themselves as well as the other group members (including the confederate) on nine-point descriptive scales regarding the statements: (*a*) This person's ideas with reference to the dot-estimate task were of the very highest quality. (*b*) This person is extremely pleasant and contributed a great deal to making this session enjoyable. (*c*) This person exerted a very great influence throughout the meeting.

RESULTS

The results regarding the newcomer's influence (assimilation) are summarized in the first three rows of Table 1. In the analyses involving "group decision error" and "estimate change" (Rows 1 and 2), a 2 × 2 factorial design was employed

TABLE 1
MEASURES OF INFLUENCE AND ACCEPTANCE OF NEWCOMER UNDER CONDITIONS OF GROUP SUCCESS AND FAILURE

	Success			Failure					Significance Tests
Group decision error	269.6			99.2					
Change in estimate	42.9			58.0					($p < .05$)
Reported change	4.2			4.9					(p about .05)
	Recipient:			Recipient:					
	New-comer (N)	Mem-ber (M)	Total	New-comer (N)	Mem-ber (M)	Total	New-comer Total	Mem-ber Total	
Ratings of "ideas"	6.3	7.1	6.7	7.2	6.1	6.6	6.7	6.6	($N \times M$, $p < .05$)
Ratings of "pleasantness"	6.8	7.6	7.2	7.4	7.7	7.5	7.1	7.6	(N vs M, $p < .05$)
Ratings of "influence"	7.0	7.3	7.1	8.1	6.4	7.3	7.6	6.8	($N \times M$, $p < .05$) (N vs M, $p < .05$)
Percentage of notes sent	34.4	37.8	36.1	33.3	36.6	34.9	33.9	37.2	(N vs M, $p < .05$)

Note.—The measures involving a percentage, "change in estimate" and "percentage of notes sent," were converted by an arc sine transformation.

including sex as a main effect. The decisions submitted by the failure groups tended to be more accurate (more in agreement with the newcomer) than the decisions of successful groups; and the failure groups tended to change their estimates more in accordance with the newcomer's estimate. However, only the latter results were statistically significant ($p < .05$).

Regarding the phenomenological measure of influence (Row 3), since the members' initial estimates were of varying degrees of accuracy, it was considered expedient to control for this variation by classifying the members of each group according to the relative accuracy of their initial estimates. The analysis of variance employed follows a technique described by Henderson (1959) for combined nested classifications (analyses involving a comparison of individuals within the same group) and factorial classifications. Members of successful groups reported less change of opinion than members of failure-groups (p about .05). Moreover, the member whose initial estimate was least accurate tended to report greater change ($p < .05$). Thus, it may be presumed that the perceived change was in the direction of the newcomer's position.

In regard to the rating items, the newcomer and a regular group member were compared under all experimental conditions. The regular member selected for comparison in each group was the person whose initial estimate was most nearly accurate or similar to the newcomer's estimate. Thus, the ratings by the members medium and low in accuracy of the newcomer and most accurate naive member were analyzed using the same complex analysis of variance technique. (The results were not appreciably changed when ratings of either of the naive members medium or low in accuracy were used as controls rather than those of the most accurate naive member.) Interaction between success-failure and newcomer-member status was significant (Row 4, $p <$ 05). Under the success condition, the regular member was rated higher than the knowledgeable newcomer with reference to quality of ideas; but under the failure condition, the newcomer was rated higher than the regular member. Similar results were obtained concerning ratings of pleasantness (Row 5) and influence (Row 6); but only the latter results were statistically significant ($p <$ 05). Moreover, the regular member was rated higher than the newcomer ($p <$ 05) with regard to "pleasantness" but with regard to "influence" the newcomer was rated higher than the regular member ($p < .05$).

Finally, the regular member received more notes than the newcomer from the remaining two naive members (Row 7, $p < .05$).

DISCUSSION

The results appear to include three major findings. First, the knowledgeable newcomer tended to be more effective or influencial in the failure groups than in the successful groups. Moreover, the newcomer tended to be rated higher than matched regular members under conditions of failure; but under success conditions the results were reversed Finally, in general, the matched regular member tended to be rated higher than the knowledgeable newcomer.

Clearly, the first two findings are mutually supporting and are in accord with the initial hypothesis: the newcomer is accepted to a greater degree by groups with a history of failure than by the successful groups. In accordance with our initial assumption, members of failure groups may perceive the newcomer as a possible resource and a possible stimulus to needed change, while successful groups may tend to perceive him as a restriction or supernumerary and a stimulus to potentially disruptive change. Since degree of assimilation was inferred from indices of influence, the results may indicate greater conformity in low prestige groups. Interpreted in these terms, the results of the present experiment are supported by an earlier experiment reporting greater conformity to a partner after previous failure on a task (Mausner & Bloch, 1957).

A recent experiment by Mills (1957) may be interpreted as supporting in part the former interpretation of the present findings. Mills found that when original members of the host group accepted one another, the value of the newcomer diminished. If members of successful groups tend to accept each other more than do members of failure groups (Deutsch, 1959), the findings of the two studies are consistent.

We also find that the newcomer with the correct answer was regarded as less pleasant, but more influential than the original members, regardless of their relative accuracy. Moreover, relatively fewer notes were addressed to the newcomer. These results may be attributed, in part, to the prescribed role of the newcomer. While the newcomer was quick to offer "to go along with the group," still, a degree of righteousness was inherent in the role. It may be conjectured that the naive members reacted negatively to this incongruent, newcomer role component. Variations in the roles of the newcomers are suggested for further research.

SUMMARY

It was proposed that a new member may be perceived as either a resource or as an unwelcome source of disruption, depending upon the needs of the group. Moreover, it was assumed that members of a group with a history of failure perceive it as incomplete and wanting in resources. Thus, it was hypothesized that a newcomer is accepted more readily by a group with a history of failure than by a successful group. The experiment tested this hypothesis for the special case of a newcomer equipped with the solution to a problem otherwise too difficult for the group to solve accurately. The experiment tested this hypothesis for the special case of a newcomer equipped with the solution to a problem otherwise too difficult for the group to solve accurately. The experiment was conducted in a laboratory, note-passing setting and involved a newcomer who was, in fact, a confederate with the correct answer to the critical task and a method of arriving at that answer. Groups with a failure induction were persuaded by the newcomer to a greater extent than groups with a failure induction, and rated the newcomer higher than the regular group members.

REFERENCES

BRUNNER, J. S., & POSTMAN, L. Perception, cognition, and behavior. In J. S. Bruner and D. Kretch (Eds.), *Perception and personality*. Durham, N. C.: Duke Univer. Press, 1949.

DEUTSCH, M. Some factors affecting membership motivation and achievement motivation in a group. *Hum. Relat.*, 1958, **11**, 81–95.

HENDERSON, C. R. Analysis of variance in complex balanced multiple classifications. Unpublished report, Cornell Univer., 1959.

MAUSNER, B., & BLOCH, B. L. A study of the variables affecting social interaction. *J. abnorm. soc. Psychol.*, 1957, **54**, 250–256.

MILLS, T. *Group structure and the newcomer*. Oslo, Norway: Oslo Univer. Press, 1957.

ZILLER, R. C., BEHRINGER, R., & JANSEN, MATHILDA J. The newcomer in open and closed groups. Unpublished report, Fels Group Dynamics Center, Univer. of Delaware, 1959.

(Early publication received June 19, 1959)

PSYCHOLOGY AND REHABILITATION

Beatrice A. Wright, Ph.D.
Editor

Based on the proceedings of an Institute on the roles of psychology and psychologists in rehabilitation, sponsored by the American Psychological Association and supported by a grant from the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

The volume examines the principles and assumptions underlying the modern philosophy of rehabilitation, paying special attention to the roles and functions of psychologists as members of an interdisciplinary team. Problems of training are explored in detail and in the context of the changing nature of the field. Areas of needed research are highlighted.

Price $1.50

Order from

American Psychological Association, Inc.

Dept. PAR
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SYSTEMATIC CHANGES IN WORD ASSOCIATION NORMS: 1910–1952[1]

JAMES J. JENKINS
Center for Advanced Study in the Behavioral Sciences

AND WALLACE A. RUSSELL
University of Minnesota

A NORM may be defined as a standard and a standard, as everyone knows, is an unchanging rule to which appeal may be made for precise measurement. Perhaps it is this verbal legerdemain which leads psychologists to overlook the fact that norms which are based on consensus or popularity of responses are especially susceptible to change and require frequent checking in our rapidly changing society.

Word association norms, an example par excellence of the consensus norms, have been commonly assumed to be stable and highly consistent from one time to another. Indeed, investigators rarely raise the issue of the possibility of normative changes but concentrate their attention on differences between groups already known to be disparate. Thus, in addition to the very numerous attempts to compare and contrast clinical groups, many studies have been directed toward comparisons of responses of normal individuals known to differ in some respect. Esper (1918), for example, compared the most frequent responses of American and German students to a standard list of stimulus words and Rosenzweig (1957) contrasted the responses of American and French students; children's responses have been contrasted with adult's responses (Rosanoff & Rosanoff, 1913; Woodrow & Lowell, 1916); the responses of men have been contrasted with those of women (Terman & Miles, 1936; Goodenough, 1942); and the responses of one preprofessional group have been compared with those of another (Foley & MacMillan, 1943).

While these studies have been both interesting and informative, it is illustrative of psychologists' blindness to the normative problems resulting from social change, that little attention has been given to detailed comparisons of even roughly matched groups tested in different years widely separated in time. As far as we have been able to discover only three investigations have mentioned the question.

O'Connor (1934) in the course of an extensive item analysis of the Kent-Rosanoff test data he had collected on a sample of "male factory workers" briefly compared the popular responses to those of the "mixed adult sample" of Kent and Rosanoff (1910). However, he did not pursue the matter further than remarking: "The common response to a stimulus word is, therefore, a scientific reality in the sense that it is rediscoverable by different workers in a new laboratory after a lapse of 15 years," and "the identity of a common response is more reproducible than its frequency of appearance" (O'Connor, 1934, pp. 200–201). Recently, Tresselt and Leeds (1955) compared word association responses collected in 1952 on a small sample, heterogeneous with respect to residence and occupation with "the original norm on the Kent-Rosanoff established in 1927" (actually the 1910 norms). They found a marked increase in primary responses which were responses of opposites (Dark–Light, Black–White) and coexisting pairs (King–Queen, Salt–Pepper). Finally, Dörken (1956) studied 10 high frequency stimulus-response pairs from the Kent-Rosanoff list in various normative collections and discovered that the frequency of these responses seemed to be increasing steadily since 1910.

---

[1] This study is part of a program of research in verbal behavior being conducted at the University of Minnesota. This study was sponsored by the Office of Naval Research (Contract N 8-onr-66216) under its policy of encouraging basic research. Research assistants who were especially concerned with the material were Evalyn Segal, Walter Mink, and Rochelle Johnson. The Graduate School of the University of Minnesota also assisted with research funds. The senior author is happy to acknowledge his indebtedness to the Social Science Research Council for a Faculty Research Fellowship during the period in which the analysis was conducted.

Apparently, though these writers suggest that changes are underway, a detailed study of changes in word association responses over time remained to be done.

A few years ago the writers fortuitously discovered that the 25-year-old word association norms for students at the University of Minnesota did not adequately represent the associative habits of the present student population. To facilitate the derivation of functional relationships in another problem (Jenkins, Mink, & Russell, 1958) new norms were gathered for the Kent-Rosanoff Word Association Test (K-R) for students in introductory psychology at this university. As these norms were studied and compared with the earlier Minnesota norms collected by Schellenberg (1930), it became obvious that the changes were not minor, trivial adjustments of a few response frequencies (as we had first suspected) but represented a fairly major and systematic change in the norms. Further work suggested that the change was not a peculiarity of sampling or technique but a phenomenon of such scope that a comparison with more diverse populations was indicated. Accordingly, other major American normative collections were studied.

This paper presents first the comparison between the early and later Minnesota norms and then goes on to report the major findings of the comparison of these two sets of norms with the other large-scale normative collections we have been able to discover. The purpose of the study is to attempt to make clear both the magnitude and the nature of the changes revealed by the comparisons of word association norms and to argue that these changes reveal a systematic and important trend in responding to free association tests, which deserves further careful consideration by students of verbal behavior, language and culture.

CHARACTERISTICS OF THE FIVE NORMS

We shall set down at once what is known concerning date of collection, characteristics of samples, and conditions of testing for each of the norms. All five studies employed the same stimulus list and presumably used the same order of presentation.

Norm 1910: Norms collected by Kent and Rosanoff (1910); published in 1910; collection dates not reported.

Testing: Oral form; individual test.

Sample: 1000 "mixed adults."

> Among these subjects were persons of both sexes and of ages ranging from eight years to over eighty years, persons following different occupations, possessing various degrees of mental capacity and education, and living in widely separated localities. Many were from Ireland and some of these had but recently arrived in this country; others were from different parts of Europe, but all were able to speak English with at least fair fluency. Over two hundred of the subjects, including a few university professors and other highly practiced observers, were professional men and women or college students. About five hundred were employed in one or another of the New York State hospitals for the insane, either as nurses and attendants or as workers at various trades, the majority of these were persons of common school education, but the group includes also, on the one hand a considerable number of high school graduates; and on the other hand, a few laborers who were almost or wholly illiterate. Nearly one hundred and fifty of the subjects were boys and girls of high school age, pupils of the Ethical Culture School, New York City. The remaining subjects form a miscellaneous group, consisting largely of clerks and farmers (pp. 38–39).

Norm 1925: Norms collected by O'Connor (1928, 1934); published in 1928; collected in 1925.

Testing: Presumably oral form, individual test; no details are given.

Sample: "1000 adult male factory workers"; no further description of sample.

Norm 1927: Norms collected by Schellenberg (1930); published in 1930, collected in fall, 1927.

Testing: Group test, printed form, written answers; instructions suggest time pressure. Other stimulus words were employed following the K-R list.

Sample: 925 students entering the University of Minnesota. Seventy-nine per cent were freshmen; 21% advanced students. Fifty-seven per cent men; 43% women.

Norm 1933: Norms collected by Keene (1951); published in 1951; collected in academic year 1933–34.

Testing: Oral form; individual test. Other stimulus words were employed following the K-R list.

Sample:

> Five hundred Stanford students, 276 men and 224 women . . . The group included individuals from all four undergraduate levels and also included some graduate students. They were students from the various psychology classes and their friends (p. 43).

Norm 1952: Norms collected by Russell and Jenkins (1954); published in 1954; collected in academic year 1952–53.

Testing: Group test, printed form, written answers.

Instructions suggest a time pressure. (Same instruction with regard to time as Norm 1927.)

Sample: 1,008 students in introductory psychology at the University of Minnesota. Largely sophomore population. About 60% male, 40% female.

Comparison of the 1927 and 1952 Minnesota Norms

Since the comparison of old and new Minnesota norms stimulated the remainder of the study and since this is the most readily justified as to comparability of subjects (*S*s) and method, the Minnesota comparison is given first and in detail. The 1927 and 1952 norms were compared with respect to the following characteristics: (*a*) frequency of primary, secondary, and tertiary responses, (*b*) number of changes in primary, secondary, and tertiary responses, (*c*) relation of the strength of the primary frequencies to the same stimulus words in the two norm sets, the response rank-frequency distribution and the characteristics of the changed response words.

Frequency of Primary, Secondary, and Tertiary Responses

Table 1 gives the average frequency of the primary, secondary, and tertiary responses in each of the norms as well as the total frequency of the three ranks combined. It is clear that a marked increase in the frequency of the most popular response has occurred. The primary response has *on the average* increased in strength by almost 30% of its earlier value; while the most popular response was given by 289 students per thousand in 1927, in 1952 it was given by 375 per thousand. This increase is also manifest at the second ranking response, but the difference has almost disappeared by the third response. Overall, the first three responses in 1952 accounted for the responses of 591 students per thousand, while in 1927 they accounted for only 489 responses per thousand. The change toward more popular responses has obviously been very pronounced.

Changes in Popular Response Words

The two sets of norms were compared as to the identity of particular response words. Words ranking first, second, and third in response to a given stimulus in 1927 were compared to the words ranking first, second, and third in 1952. The results are shown in Table 2.

TABLE 1

Average Frequency of the Three Most Popular Responses on the 1927 and 1952 Minnesota Norms

Response Rank	Frequency 1927 ($N = 925$)	Frequency 1952 ($N = 1008$)	Percentage 1927	Percentage 1952
1st	267.5	377.5	28.9	37.5
2nd	111.9	137.1	12.1	13.6
3rd	72.8	81.2	7.9	8.1
Total 1st, 2nd, 3rd	452.2	595.8	48.9	59.1

TABLE 2

Comparison of the Response Words Occurring in the First Three Ranks in the 1927 and 1952 Minnesota Norms

1952 Norms	1927 Norms Rank 1	2	3	Lower	Total
Rank 1	71	17	5	7	100
2	14	36	23	27	100
3	8	14	21	57	100
Lower	7	33	51		91
Total	100	100	100	91	

This table shows, as might have been expected, that the words which were primary responses in 1927 tend to be the primary responses in 1952; 71 of the 100 primary responses were identical. Fourteen of the 1927 primaries slipped to the second rank in 1952, eight to the third rank, and seven to ranks below that. The 1927 secondary responses were somewhat unstable; 17 moved up to become 1952 primaries, 14 fell to become tertiaries, and 33 fell to lower ranks; only 36 remained the same. The 1927 tertiaries have an even greater dispersion, as the table shows. Overall, 91 new words entered the top three ranks, and, of course, 91 older words moved out. It is clear that most of this turnover took place at the third rank, with the ranks becoming increasingly more stable as the strength of response increased.

Strength of Primary Responses to Particular Stimuli

In the face of the facts that some primary responses had changed and that primaries in general had greatly increased in strength, it was of interest to enquire whether the relative strength of a primary response to a given stimulus in 1927 was related to the strength of the primary response to the same stimulus in 1952. Accordingly, three correlations were calculated: (*a*) the correlation between frequency of the primary responses for *all* stimulus words on the two occasions, (*b*) the same correlation for the 71 stimulus-response pairs which had not changed, and (*c*) the correlation for the 29 pairs for which the response had changed. The over-all correlation ($N = 100$ pairs) was $+ .70$; the correlation for same-response pairs was $+.65$, and the correlation for pairs in which the response had changed was $+.48$. The three correlations are all significantly different from zero. It is clear that the relative ranks of the strength of primary responses to particular stimuli remain fairly stable over time. Inspection of the scatterplots shows that the primary responses with high strength in the 1927 norms did not tend to change but simply increased in strength as popular responding increased. The primary responses which changed tended to be those which were at relatively low strength in 1927. Such responses were displaced by close competitors in the 1927 norms which moved up to become new low strength primaries.

The Total Response Distribution

The overall distribution of frequencies of responses when plotted against response rank closely approached a Zipf-type distribution. When for the 1952 norms the average frequency of each response rank was plotted against the rank on double logarithm paper, the result was almost a straight line from the first to the 90th rank. After the 90th rank (at which point the average frequency was slightly less than one, indicating that at least some of the response hierarchies had already been exhausted) the curve fell away rapidly toward zero. This distribution may be compared to the rank-frequency distribution for the Kent-Rosanoff 1910 norms reported by Skinner (1937) and the rank-frequency distribution for 200 "other" stimulus words (which were *not* the K-R stimuli) collected by Schellenberg (1930) and plotted by Cook and Skinner (1939). The present distribution is appreciably more linear than either of the earlier distributions and is somewhat steeper, having a slope of approximately -1.4 as contrasted with roughly -1.3 for the Kent-Rosanoff distribution and -1.1 for the Schellenberg "other" words.

The Nature of the Changed S-R Pairs

The stimulus words which evoked differing primary responses in 1952 and 1927 are listed in Table 3 with their old and new responses. The strength of each response is indicated by the percentage of *S*s making that response.

The writers are well aware of the difficulties of classifying associative pairs and are especially wary of "logical" as opposed to "behavioral" classifications. Nevertheless it is clear that at least nine of the 29 stimulus-response pairs in this group that were popular in 1927 were *superordinate* pairs (i.e., Butter-

TABLE 3

STIMULI AND THEIR PRIMARY RESPONSES FOR PAIRS WHICH WERE DIFFERENT IN THE 1927 AND 1952 NORMS

Stimulus	1927 Response	Percentage of Subjects Making the Response	1952 Response	Percentage of Subjects Making the Response
Sickness	Illness	19	Health	37
Mutton	Sheep	34	Lamb	36
Comfort	Ease	13	Chair	12
Short	Long	33	Tall	39
Butterfly	Insect	20	Moth	14
Whistle	Noise	17	Stop	13
Wish	Desire	20	Want	12
Beautiful	Pretty	19	Ugly	21
Window	Glass	26	Door	19
Citizen	Man	21	U.S.	11
Red	Color	23	White	22
Sleep	Rest	25	Bed	24
Working	Labor	11	Hard	13
Earth	Ground	17	Round	13
Trouble	Sorrow	8	Bad	9
Cabbage	Vegetable	40	Head	16
Yellow	Color	29	Blue	15
Justice	Judge	14	Peace	25
Bible	Book	29	God	23
Sheep	Animal	18	Wool	20
Bath	Water	23	Clean	31
Blue	Color	24	Sky	17
Stove	Heat	25	Hot	23
Doctor	Sick	15	Nurse	24
Thief	Robber	22	Steal	28
Lion	Animal	33	Tiger	26
Joy	Happiness	16	Happy	21
Baby	Child	18	Boy	16
Cheese	Food	12	Crackers	11

fly-Insect, Red-Color, Cabbage-Vegetable, Yellow-Color, Bible-Book, Sheep-Animal, Blue-Color, Lion-Animal, and Cheese-Food). If one permits a little ambiguity, the number of pairs classified as superordinate increases (i.e., one may add Whistle-Noise, Wish-Desire, Citizen-Man, etc.) The 1952 responses stand in sharp contrast to this. *None* of the new responses is clearly of a superordinate nature, and only two could be considered to be even marginal superordinates (Sickness-Health? Butterfly-Moth?). It appears from these data that superordinate or classifying responses have decreased in popularity since 1927.

The second major characteristic of the 1927 responses seems to be *synonymity*, but it is difficult to obtain reliable judgments of this relation which are separate from and independent of superordination. The reader may judge for himself how nearly synonymous the stimuli and responses are.

The prevailing characteristics of the 1952 responses seem to be, on the one hand, "coordination" and, on the other hand, "completion" or some form of *sequential relation*. Frequently, these "types" are fused in the new stimulus-response relation. As an example, Cheese–Crackers may be thought of as a pair of coordinates within a class of edibles or as a highly frequent sequential combination in everyday language, "cheese and crackers." The same is true of Lion–Tiger, Doctor–Nurse, Justice–Peace, Red–White, Sickness–Health, etc. Again, systems of classification are inadequate to express the difference between the norms quantitatively.

Overall, the changes in primary responses seem to suggest that current responses are more specific, more concrete, and perhaps more determined by high frequency verbal sequences than the older responses. The older responses which have been discarded tended somewhat more toward abstraction, with the most popular responses being those of superordination and synonymity.

Discussion

At this point the import of the contrast of the 1927 and 1952 norms needs to be stressed. We had suspected from experimental evidence that the word association norms had changed, but we had not been aware of the nature of the change. Before the data were gathered, we had assumed that the 1952 sample would show less agreement and greater diversity in responses than the 1927 sample. The 1952 sample was more heterogeneous (a higher proportion of youth were in school; veterans on the GI Bill represented persons with diverse service backgrounds and in many cases individuals who would not otherwise have been in school, etc.); and the sample was one year advanced in education over the 1927 group (which is supposed to make for more individual responding). In short, we anticipated exactly the opposite findings from those we discovered. Instead of being more heterogeneous, in word association the group was startlingly more homogeneous. There had been marked change, but it was almost all toward conformity.

Clearly there was some stability in free association data across the 25-year period. The high frequency responses tended to persist as the most popular responses, and the ranks of the most popular responses tended to remain the same. On the other hand, we had found a greatly increased tendency on the part of the 1952 *S*s to use the popular responses and a marked tendency to change low frequency primaries of the superordinate sort to other primary responses which tended to be more specific coordinates or completions.

The literature offered little aid in explaining our data. Kent and Rosanoff (1910) had pointed out that their highly educated *S*s made *less* use of the popular response words than the less well educated *S*s in the norm group. Cook and Skinner (1939) had used this finding to explain the distribution of responses on Schellenberg's (1930) 200 "other" stimulus words, where *S*s had in fact used the popular responses less often and the infrequent responses somewhat more often than Kent and Rosanoff's *S*s had for comparable ranks on different stimulus words. But Cook and Skinner were apparently not aware of the fact that on the 100 K-R stimulus words where data were in fact comparable, Schellenberg's sample used the most popular responses *more often* than Kent and Rosanoff's "mixed adults."

We may at this point oppose Kent and Rosanoff's finding and argue that increased education increases the use of popular responses, thus accounting for the differences between the 1927 and 1910 norms and attributing the difference between the new and the old Minnesota norms to the extra year of education, or we may argue that 1952 sophomores were actually less well educated than 1927 freshmen (and ignore the contrast between the 1927 and 1910 norms) or, finally, we may attribute the difference to something other than educational level such as the methods employed in the various studies or the time-lapse and cultural change between samplings. The writers feel that the last alternative is the only one worthy of serious consideration.

If one were going to attribute the difference to the method employed, one would presumably argue that the group-written form produces more common associates than the individual-oral form, thus accounting for the difference between the 1927 and 1910 norms, and then feel free to introduce some other sort of variable to explain the difference between the 1927 and 1952 Minnesota groups.

Woodworth and Schlosberg (1954), however, feel that the obvious time-pressure of the oral-individual form of the test leads to even greater use of the popular associates than the group-written method. In their chapter on "Association" they compare the results of individual testing of Brown University students with the 1929 Minnesota norms and suggest that the great increase in the frequency of popular responses is a consequence of the difference in method. Inspection of the limited data they present, however, shows a striking correspondence to the 1952 norms, which suggests that the method itself may make only a secondary contribution to the high frequencies observed.

An unpublished study by Clousing (1927) seems to confirm the negligible effect of method (as long as both administrations suggest time pressure). In her study, no differences were found between the group-written and the individual-oral administrations in the frequency of primary, secondary, or tertiary responses. Keene (1951) asserts that there is little difference between the two modes of testing, although he does not present his group data for comparison with his tables of individual data. From the data currently available, explanation of the differences between norms in terms of the differences between individual and group testing does not seem satisfactory.

At this point it was decided provisionally to accept the findings concerning the Minnesota norms as indicating general trends in associative behavior, independent of specific sample or methodological differences. We proposed to test the findings by examining them in the larger context of the 1910, 1925, and 1933 norms. We felt that three hypotheses were testable.

1. Responses of *S*s to the K-R stimuli are increasingly concentrated among the popular responses for all United States adult samples. We proposed to examine the five norms to see if there was evidence for a steady increase in the frequency of the common responses.

2. Words used as responses to the K-R stimuli tend to change slowly over time, with the highest ranking responses having the highest stability. The amount of change for any particular rank should be a function of the time elapsing between the collections of the norm data.

3. "Abstract" responses to stimuli tend to decrease in popularity across the time period encompassed by this study. Specifically, if superordinates are taken as representative of abstractions, they will be found to occur less and less often as responses to stimuli, as the date of the norms becomes more nearly current.

It should be noted that these hypotheses vary in the amount of support they receive directly from the Minnesota comparison already made. The first hypothesis has the most direct support, since the recent Minnesota norms had already been found to be more concentrated in the popular associates than the earlier norms, which in turn were known to be more concentrated than the Kent-Rosanoff. Only if the O'Connor and Keene norms deviated from the trend could the hypothesis be denied. The second hypothesis, however, has many opportunities for confirmation or denial. Each set of norm data may be used as a standard, and the amount of difference between the standard set and other sets may be directly calculated. The third hypothesis similarly has several opportunities for real tests and may be relatively independent of the original data which suggested it (the data on changes in primary responses alone). Here it was decided to obtain a relatively objective classification scheme, classify all superordinates for the stimulus words, and examine their frequency of occurrence in the several norms. The prediction, of course, was that a decrease in their use would be found over time.

RESULTS OF GENERAL NORM COMPARISON

Frequency of Popular Responses

Table 4 lists the frequences of the most popular responses (first, second, and third rank) for all five norm groups. It is immediately clear that the simple hypothesis of an increase in response frequencies over time will not by itself account for the data. The two norms which are closest together in time, 1925 and 1927, are widely deviate with respect to response frequencies. The 1925 norm is very similar in frequency to the 1933 norm and more like the 1952 norm than it is like the 1910 norm. The 1927 norm is most like the 1910 norm. The similarities and differences here are remarkable in that they cut across the types of samples and methods of administering the tests.

It is tempting to suggest that two factors play a part in the variation found in our data. One factor would be the method of test administration (accepting the Woodworth and Schlosberg [1954] suggestion that oral administration leads to higher popular response frequencies); the other factor would be the general cultural shift which progressively increases the use of common responses. At any one time, one would expect the oral method to yield higher popular response frequencies than the group-written method, but over time both methods would show the increasing drift toward popular responses. The data for primary response frequencies show parallel lines of increase when plotted for oral-individual and group-written administrations separately.

TABLE 4

AVERAGE FREQUENCY OF THE THREE MOST POPULAR RESPONSES TO 100 STIMULI IN THE FIVE NORMATIVE STUDIES

Response Rank	1910[a]	1925[a]	1927[b]	1533[b]	1952[b]
First	260	338	289	357	375
Second	122	134	121	125	136
Third	77	80	79	78	80
Total 1, 2, 3	459	552	489	560	591

[a] Singular and plural forms of responses were combined in computing frequencies to make these comparable to the other norms.

[b] All frequencies expressed as frequencies per 1000.

In view of the fact that there is no experimental support for the first factor and that so little is known concerning the actual collection of the 1925 norms, this explanation must be left as a conjecture and as a suggestion and stimulus for further research.

An empirical finding which is quite clear, however, is that at least college populations currently show a very high usage of common responses. The data for Brown University students has already been mentioned as reflecting the high frequencies found in the 1952 and 1933 norms. In addition, support for this assertion is found in the normative compilation of Meals, Herrick, and Merow (1948) at the University of Pennsylvania for 50 of the K-R words. These investigators apparently selected stimuli which elicited high response frequencies as indicated by the 1910 or 1927 norms. Frequency data for the same fifty stimulus words were abstracted from the 1952 norms for comparison with the Pennsylvania data. The proportions responding at each rank are paired for the Pennsylvania and Minnesota norms respectively: first rank, .52, .47; second, .11, .13; and third, .07, .07. It is easily seen that the same intense concentration on popular responses is found for Pennsylvania college students.

Overall, with respect to the first hypothesis, it may be said that the use of popular responses in the free association test is currently at a very high level in college populations, whether they are tested with the individual-oral or group-written method The data strongly suggest that the frequency of popular responses is increasing over time, but that other factors in addition to the general trend are important in determining popular response frequencies.

Identity of Words Used as Responses

Table 5 lists the percentage of identical responses at each rank level for each pairing of the five normative groups. For the primary responses it is clear that the hypothesis that there is a steady change with time is acceptable. Each norm group shows most overlap with the norm group which is closest chronologically.

The primary responses in the 1952 norms are identical with 73 of the primary responses in the 1933 norms, 71 of the primaries in the 1927 norms, 65 of the primaries in the 1925 norms, and only 58 of the primaries in the 1910 norms. Conversely, the 1910 norms show identical primaries with 81 responses in the 1925 norms, 79 responses in the 1927 norms, 74 responses in the 1933 norms, and 58 responses in the 1952 norms. Over the entire table in similar fashion all the values are in accord with the general hypothesis.

For the secondary and tertiary responses, similar results are found with good conformance to the hypothesis except that the 1925 and 1910 norms are most alike in both cases and the 1933 norms are slightly less similar to the 1927 norms than would have been expected. From either extreme, 1910 or 1952, the progression is quite orderly, however.

While the 1948 Pennsylvania norms referred to above are probably somewhat biased by the selection of stimuli known to give high frequency response, it is interesting to note in passing that they show very high agreement with the 1952 norms. Eighty-two per cent of the primary responses are identical, 42% of the secondary responses are identical, and 32% of the tertiary responses are identical. Inspection of Table 5 reveals that these approximate very closely the percentages of overlap of the 1925 and 1927 norms (83, 48, and 32%) which are separated by a similar short time period.

TABLE 5

PROPORTION OF IDENTICAL RESPONSES IN THE FIVE NORMATIVE STUDIES

Norm	Primary Responses				
	1910	1925	1927	1933	1952
1910	—	81	79	74	58
1925	81	—	83	79	65
1927	79	83	—	80	71
1933	74	79	80	—	73
1952	58	65	71	73	—

Norm	Secondary Responses				
	1910	1925	1927	1933	1952
1910	—	53	44	26	26
1925	53	—	48	47	34
1927	44	48	—	43	36
1933	26	47	43	—	46
1952	26	34	36	46	—

Norm	Tertiary Responses				
	1910	1925	1927	1933	1952
1910	—	36	20	19	10
1925	36	—	32	29	19
1927	20	32	—	27	21
1933	19	29	27	—	29
1952	10	19	21	29	—

TABLE 6

NUMBER OF WORDS DIFFERENT IN THE FIRST THREE RESPONSES
(300 words)

Norm	1910	1925	1927	1933	1952
1910	—	69	79	94	123
1925	69	—	66	71	95
1927	79	66	—	88	91
1933	94	71	88	—	77
1952	123	95	91	77	—

Table 6 presents the other side of the consistency picture, showing how many *different* words are found in the first 300 response words for each of the norm pairs. These data conform closely to the hypothesis except in the degree of difference between the 1933 and 1927 norms. In all other respects the table is in accord with the hypothesis.

Considering all the findings, it seems clear that a major determinant of response-overlap or response-identity in word association norms is the time interval between collections of data. Where deviations from this simple prediction pattern do appear, there is a suggestion that the norms derived by the oral-individual method are slightly more alike than would be expected from time lapse prediction. That other variables are probably active is of course conceded, but it is stressed here that time lapse alone accounts for most of the systematic variation.

Use of Superordinates

In order to obtain an unbiased and relatively objective standard for deciding which responses were superordinates, a written test was devised. The test consisted of a set of 100 sentences of the form "—— is a member of the class of ——." Each sentence began with one of the Kent-Rosanoff stimulus words. This

TABLE 7

WORDS DEFINED AS SUPERORDINATES BY THE CRITERION OF THE SUPERORDINATE TEST

Stimulus	Superordinate Response	Frequency ($N = 29$)	Stimulus	Superordinate Response	Frequency ($N = 29$)
Red	Color	29	Swift	Speed	21
Yellow	Color	29	Spider	Insect	21
Bread	Food	29	Bed	Furniture	20
White	Color	29	Ocean	Water	20
Green	Color	29	Slow	Speed	20
Butter	Food	28	Scissors	Tool	19
Blue	Color	28	Butterfly	Insect	19
Black	Color	28	Heavy	Weight	18
Lion	Animal	27	Bible	Book	18
Sour	Taste	27	Cabbage	Vegetable	18
Sheep	Animal	26	Blossom	Flower	17
Cheese	Food	25	Sickness	Health	17
Hammer	Tool	25	Music	Art	17
Eagle	Bird	25	Quiet	Sound	16
Chair	Furniture	25	Lamp	Furniture	16
Table	Furniture	25	River	Water	16
Bitter	Taste	23	Salt	Food	15
Sweet	Taste	23	Stomach	Organ	15
Fruit	Food	22	Anger	Emotion	15
Mutton	Meat	22			

test was given to 29 students in an introductory psychology laboratory course. A superordinate response was defined as any sentence completion that was given by 15 or more of the students taking the test. We realize that this procedure is arbitrary in the extreme, but it includes the stimulus-response pairs which most judges would agree are "really" superordinates, provides an unequivocal criterion for counting responses, is unbiased with respect to the normative data to be judged, and excludes words for which superordinates are, to say the least, difficult to define. This procedure yielded agreement as to superordinate for 39 of the 100 K-R stimuli. These stimuli and their superordinates are listed in Table 7 with the frequency with which the superordinate was given.

The occurrences of the superordinate responses to appropriate stimuli were counted in the 1910, 1925, and 1952 norms where complete data were available. The 1927 and 1933 norms could not be used in this comparison because only the three most popular responses are available for each stimulus. Accordingly, a separate tabulation was made of the 26 superordinate responses which were present in the 1927 norms and the 18 superordinate responses which were present in the 1933 norms. The data are presented in Table 8.

Inspection of the table verifies the hypothesis concerning the decline of superordinate responses. The frequency of occurrence of the 39 superordinate responses identified has declined in the 1952 norms to just slightly more than half the frequency manifested in the 1910 norms. The frequency of occurrence of the 26 superordinates present in the 1927 norms (1st, 2nd, and 3rd ranks), even though biased upward for that norm set because of selection, shows a marked decrease from 1910 to 1952. The frequency of occurrence of the 18 superordinates present in the 1933 norms does not even appreciably reflect the selective bias but falls steadily from 1910 to 1952.

A detailed response-by-response comparison confirms this trend as a general case not attributable to a marked change in a few responses. Comparison of the 1910 and 1952 norms shows decreasing frequencies for 33 of the 39 superordinate responses. Further, for what might be called the clearest cases of superordination (those identified on the superordinate test with a frequency of 20 or greater), 24 out of 25 responses have declined. Comparison of the 1925 and 1952 norms shows a decline in 31 of the 39 superordinate responses and 23 of the 25 "strong" superordinate responses. The comparison of the 1910 and 1925 responses, while not as overwhelming, is still conclusive; 30 of the 39 superordinates declined in frequency and 21 of the 25 "strong" superordinates declined in frequency. The

TABLE 8

USE OF SUPERORDINATES AS RESPONSES

	1910	1925	1927	1933	1952
39 Superordinates Average percentage	17.5	15.4	—	—	9.6
26 Superordinates[a] Average percentage	24.7	22.3	22.6	—	13.9
18 Superordinates[a] Average percentage	31.1	28.3	27.9	22.4	19.2

[a] For the 1927 and 1933 norms only the first three responses are available. Twenty-six superordinates occurred in the first three ranks in 1927 and 18 superordinates occurred in the first three ranks in 1933. All those appearing in 1933 also appeared in 1927. The second and third rows of the table are biased upward for 1927 and 1933, therefore.

decline of popularity of the superordinate response appears to be strikingly demonstrated.

Discussion

The first hypothesis concerning the general increase in the frequency of popular responses appears only partially confirmed. It is clear from the data that popular responses are currently at very high frequencies in college populations, but the five normative summaries do not support the notion of a simple trend in this direction over time on all United States adult samples. The 1925 sample shows a much greater use of popular responses than the 1927 sample. The most parsimonious explanation at this time seems to be a two-factor explanation recognizing both the drift toward popular responses over the time-span included in this study and a tendency for the oral-individual test technique to generate more popular responses than the group-written technique.

The second hypothesis concerning the change in identity of popular response words over time seems to be adequately confirmed. Norms collected closest in time show the maximum amount of overlapping in responses (regardless of sample differences), and high frequency responses prove to be much more stable than lower frequency responses. There is a further suggestion that test technique contributes to similarity of responses.

The third hypothesis concerning the decline in the use of superordinates as word association responses is likewise adequately confirmed. A clear and highly significant trend was discovered and the evidence further indicates that the more clearly the responses were identified as superordinates, the more marked the trend was seen to be.

Interpretations of these findings are at best conjectural. The writers have been tempted to discuss the study in terms of the growth of "mass culture," "other-directedness" and similar constructs much used today in characterizing our culture. The enormous growth of the mass media, the standardization of school courses, procedures and text books, and the presumably greater homogeneity of our verbal culture are appealing variables to invoke to explain the phenomena manifest in our data, but no *evidence* can now be marshalled to support such a discussion.

A much simpler explanation for much of the data suggests itself; namely, that test-taking attitudes have been changing over the time period discussed here. The work of Dunn, Bliss, and Siipola (1958), Siipola, Walker, and Kolb (1955), and to a lesser extent that of Flavell, Draguns, Feinberg, and Budin (1958) clearly establishes the importance of the time-pressure set in producing popular and "superficial" responses, such as contrasts and coordinates, and decreasing responses which have a synonymous or superordinate relation to the stimulus. These modern studies of test-taking attitude mirror the effects achieved by Wells in a much older study of practice effects in free association. Wells (1911) gave his *S*s 20 association lists of 50 words (one list per day) and then retested them on the first two lists. He observed a marked decrease in reaction time and a change to "superficial" associations. He reports the qualitative changes in response for each of his six *S*s, and for each one it is remarked that the frequency of superordinate responses decreased (the decreases vary from one-half to two-thirds of the original number of responses so classified).

It would appear that practice effects ("test-wiseness") and time pressure ("test-attitude") bring about much the same effects: so-called superficiality of response, decrease in superordinates, and increased popular responding. If we assume that the oral-individual form maximizes time pressure and that more recent dates give us more and more test-sophisticated *S*s we would seem to have all the necessary variables required to account both for the frequency data and the superordinate data.

It is true of course that "test-wiseness," or "test-docility" of *S*s is after all a general product of our culture and that this explanation may still be related to general questions of cultural change on a grand scale. However, this explanation has the virtue of greater specificity and seems to furnish a point of attack for further investigations which the more sweeping explanation fails to furnish.

It should be noted beyond this, however, that regardless of the explanation given for the increasing frequencies of popular responses and decreasing frequencies of abstract responses, the systematic change in the overlapping of specific stimulus-response pairs as a function of time-lapse between normative studies is in the main separable from these phenomena. This systematic change appears to the writers to be attributable not only to the drift away from abstract responses but also to changes in the "meaning" of words (e.g., "lamp" in 1910 apparently meant an "oil lamp" in contrast to its current meaning) and to particular recent experiences concerning the words which were not duplicated during the other norm-collection periods (e.g., in 1952, a strenuous presidential campaign was in progress and "whistle-stop" became a very popular associative pair). These changes could profitably be studied as part of a general study of the change in meaning of common words over time.

SUMMARY

A comparison was made of word association norms collected in 1927 and in 1952 for University of Minnesota students. It was found that primary responses had greatly increased in frequency and that the first three responses to the stimulus words accounted for 59% of all responses in 1952 as contrasted with 49% in 1927. Seventy-one per cent of the primary responses were identical. Primary responses were found to be more stable than secondary or tertiary responses. The frequency of the primary response to individual stimuli correlated +.70 between the two norms. Primary responses which had changed, tended to be low-strength superordinate responses which were replaced by coordinates and completions.

Three hypotheses arising from the 1927–1952 comparison were examined with respect to five major collections of free association data: the 1910 Kent-Rosanoff norms, the 1925 O'Connor norms, the 1933 Keene norms, and the 1929 and 1952 Minnesota norms. The first hypothesis, that there is a general tendency for the frequency of popular responses to increase with time, received partial confirmation. It was concluded, however, that other factors, such as testing method probably exercised some effect in addition to the time change.

The second hypothesis was that the words used as responses to stimuli tend to change slowly but systematically over time with the highest ranking responses having the highest stability. This hypothesis was confirmed.

The third hypothesis was that abstract responses (specifically superordinates) have tended to decrease in popularity across the time period of this study. Superordinates were identified for the Kent-Rosanoff stimuli by means of a sentence completion technique. The hypothesis was confirmed.

Possible explanations for these phenomena were mentioned and it was suggested that a general change in test-taking attitudes might account both for the increase in popular responses and the decrease in abstract responses. It was further suggested that the change in individual response terms (aside from the frequency and abstraction changes) might be attributable to changes in the meaning of particular stimuli over time.

REFERENCES

CLOUSING, JOSEPHINE. *The relation between personality trait ratings and free association responses.* Unpublished Master's thesis, Univer. of Minnesota, 1927.

COOK, S. W., & SKINNER, B. F. Some factors influencing the distribution of associated words. *Psychol. Rec.*, 1939, **3**, 178–184.

DÖRKEN, H., JR. Frequency of common associations. *Psychol. Rep.*, 1956, **2**, 407–408.

DUNN, SANDRA, BLISS, JOAN, & SIIPOLA, ELSA. Effects of impulsivity, introversion, and individual values upon association under free conditions. *J. Pers.*, 1958, **26**, 61–76.

ESPER, E. H. A contribution to the experimental study of analogy. *Psychol. Rev.*, 1918, **25**, 468–487.

FLAVELL, J. H., DRAGUNS, J., FEINBURG, L. D., & BUDIN, W. A microgenetic approach to word association. *J. abnorm. soc. Psychol.*, 1958, **57**, 1–7.

FOLEY, J. P., JR., & MACMILLAN, Z. L. Mediated generalization and the interpretation of verbal behavior: V "Free association" as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299–310.

GOODENOUGH, FLORENCE L. The use of free association in the objective measurement of personality. In Q. McNemar & M. A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942. Pp. 87–103.

JENKINS, J. J., MINK, W. D., & RUSSELL, W. A. Associative clustering as a function of verbal association strength. *Psychol. Rep.*, 1958, **4**, 127–136.

KEENE, C. M. Commonality of response on a word-association test: A study of standardization procedures and an attempt to forecast moderate emotional maladjustment. Unpublished doctoral dissertation, Stanford Univer., 1951.

KENT, GRACE H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, **67**, 37–96, 317–390.

MEALS, D. W., HERRICK, C. J., & MEROW, ROSALIND.

Scoring sheet for reaction word experiment. Univer. Pennsylvania, 1948. (Mimeo.)

O'CONNOR, J. *Born that Way*, Baltimore: Williams & Wilkins, 1928.

O'CONNOR, J. *Psychometrics*, Cambridge: Harvard Univer. Press, 1934.

ROSANOFF, ISABEL R., & ROSANOFF, A. J. A study of association in children. *Psychol. Rev.*, 1913, **20**, 43–89.

ROSENZWEIG, M. R. Etudes sur les associations des mots. *Année Psychol.*, 1957, No. 1, 23–32.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. USN Tech. Rep. No. 11, 1954, ONR N8 onr-66216, Univer. Minnesota, 1954.

SCHELLENBERG, P. E. A group free-association test for college students. Unpublished doctoral dissertation, Univer. of Minnesota, 1930.

SIIPOLA, ELSA, WALKER, W. NANNETTE, & KOLB, DOROTHY. Task attitudes in word association, projective, and nonprojective, *J. Pers.*, 1955, **23**, 441–459.

SKINNER, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, **1**, 71–76.

TERMAN, L. M., & MILES, CATHERINE C. *Sex and personality: Studies in masculinity and femininity*, New York: McGraw-Hill, 1936.

TRESSELT, M. E., & LEEDS, D. S. The Kent-Rosanoff word association: I. New frequencies for ages eighteen–twenty-one and a comparison with Kent-Rosanoff frequencies, *J. genet. Psychol.*, 1955, **87**, 145–148.

WELLS, F. L. Practice effects in free association. *Amer. J. Psychol.*, 1911, **22**, 1–13.

WOODROW, H., & LOWELL, F. Children's association frequency tables. *Psychol. Monogr.*, 1916, **22** (5, Whole No. 97).

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental Psychology.* (rev. ed.) New York: Holt, 1954.

(Received July 21, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

NYSTAGMUS:

AN INTEGRATIVE NEURAL DEFICIT IN SCHIZOPHRENIA[1]

WALTER W. LEACH
Mississippi Southern College

THE significance of body image in psychopathology calls attention to proprioceptor mechanisms. One of the most important is the vestibular apparatus because of its reflexive adjustment of the body's position. Through integration of impulses from the semicircular canals at cerebellar and oculomotor centers, an individual maintains equilibirum and orientation to his environment. Mowrer (1932) and Holsopple (1929) demonstrated that semicircular canal eye reflexes maintain normal visual perception during head movements. The observation that some schizophrenic patients have a marked disturbance in body image and orientation to their environment suggests disturbances in their equilibratory processes.

The semicircular canals regulate muscle tonus throughout the body. Angyal (1935, 1936, 1937) describes a schizophrenic syndrome in which a bizarre experience of the body occurs, e.g., an impression of heaviness or lightness. Angyl and Blackman (1940) used the Barany test to compare the postrotation nystagmus of 58 schizophrenic and 20 normal subjects (*S*s). They lacked a recording device and could set no criteria for the eye movements that constitute a nystagmic beat. In addition, they made no observations during stimulation. Objections can be raised concerning the nature of the stimulation in the Barany test (10 rev. in 20 sec. with a sudden stop). This extremely intense and erratic stimulus is greatly beyond appropriate physiological limits and frequently produces distressful side-effects.

The primary purpose of this paper is to compare experimentally groups of schizophrenic and normal *S*s by measuring characteristics of their nystagmus. Amplification of the finding of Angyal and Blackman (1941) required apparatus to provide controlled uniform angular acceleration and to record nystagmus during as well as after rotation. A neural correlate of schizophrenia could provide an objective basis for diagnostic classification. In its broader implications, this experiment is aimed at a unified approach to biological and psychological functioning.

METHOD[2]

Subjects. Seventy-five male inpatients, diagnosed as schizophrenic by the staff of the Lyons, New Jersey, VA Hospital, made up the experimental group. Their ages ranged from 19 to 49 (mean, 29; *SD*, 8). Patients with a record of neurological disease were excluded. Average duration of hospitalization was less than six months.

Ninety-six noninstitutionalized males, ranging from 14 to 76 years of age (mean 26; *SD*, 10), were in the normal control group. Nearly all were university students. The same basis for exclusion applied as for the patient group.

Apparatus. A rotating chair designed by J. Q. Holsopple and the writer provided controlled intensities of uniform angular acceleration. Since a falling weight accelerates at a uniform rate if meeting constant resistance, an essential requirement was that the chair be precisely balanced. During the fall of a weight attached to the axis by a cable, the chair accelerated until the cable unwound. It then decelerated at a uniform rate as the momentum lifted the weight. The total weight of the chair while rotating *S*s of varying weights was kept constant. Each *S* was weighed, and lead ingots in a pan attached to the chair were added or subtracted accordingly.

The intensity of acceleration was determined by the amount of weight attached to the cable. Uniformity and intensity were empirically calibrated at each value of the stimulus. The importance of known uniformity is seen if one considers how unsatisfactory an average value would be in other sense departments.[3]

A standard ophthalmograph (American Optical Company, Model 1520) mounted on the foot plate of the chair was used to record nystagmus during and after stimulation. Start and end of stimulation were signaled directly on the ophthalmographic film. The head rest aligned the vertical axis of the head with that of the chair to stimulate the horizontal semicircular canals.

---

[1] This paper is based on a dissertation presented to the faculty of Princeton University in partial fulfillment of the requirements for the degree of doctor of philosophy. The writer wishes to express his appreciation for assistance to E. G. Wever and the late J. Q. Holsopple.

[2] For a more detailed description see Leach, 1953.

[3] For a discussion of the problems involved cf., Holsopple, 1929; Mowrer, 1932; and ter Braak, 1936.

Procedure. After *S* was seated in the chair and the head rest was adjusted to prevent movement, he was instructed to look at an "X" a short distance from his eyes. Corneal reflections from each eye were focused onto the film. A record of eye steadiness was obtained before each rotation. *S*s were stimulated in three separate rotations with mild, moderate, and strong accelerations administered in a sequence of 2, 5, and 12° per sec. per sec. Duration of the rotation was 30 sec. at 2°, 21 sec. at 5°, and 16 sec. at 12°. The first and third rotations were counter clockwise; the second was clockwise.

A brief explanation of the purpose was given to each *S*. Cooperation of psychotic *S*s was facilitated by eliminating all cues and any need for communication. The lights of the ophthalmograph were shielded so that none of the surroundings were visible.

Measures. The following measures were made from the nystagmograms at each stimulus intensity:

1. Acceleration latency: time in seconds from beginning of rotation until the completion of the first nystagmus.
2. Deceleration latency: time in seconds from beginning of rotation until the completion of the first reversed nystagmus.
3. Number of acceleration nystagmi during rotation.
4. Number of deceleration nystagmi during rotation.
5. Number of postrotation nystagmi.
6. Duration of postrotation nystagmi: time in sec. until (*a*) no nystagmus appeared for three sec. or more, or (*b*) until response rate dropped by one-half or more. Rotation and postrotation rates were obtained by dividing the number of nystagmi for each period by the duration of each in sec.

The nystagmograms were read by passing the 35 mm. film over a translucent glass in front of a 25-watt light bulb. Since the film was exposed at the rate of ½ inch per sec., time was measured in units of .12 sec. along the edge of a ruler calibrated in sixteenths of an inch. A nystagmic beat was counted for each sequence of slow deviation of the eye tracings from the fixation point and a compensatory saccadic return. The direction of the eye movements was to either the right or left, depending upon (*a*) whether the stimulus was acceleration or deceleration, and (*b*) the direction of rotation. In a clockwise rotation, the slow phase during acceleration was to the left and the compensatory phase to the right. This sequence reversed to right-left during deceleration. In counter-clockwise rotation, these directions were exchanged.

RESULTS

The schizophrenic group showed diminished nystagmic reactivity for all of the measures except deceleration latency. Their deficits during stimulation appeared in their longer reaction times and slower rates of response. Postrotation nystagmus showed a similar reduction. The percentage deficit and the statistical significance of the differences between the schizophrenic patient and the normal control groups are given in Tables 1 and 2.

The amount of deficit varied, depending upon the intensity of stimulation. It was least to strong stimulation and greatest for weak stimulation for all measures except number of deceleration nystagmi. As the stimulus inten-

TABLE 1

COMPARISON OF ROTATION NYSTAGMUS AT 3 STIMULUS INTENSITIES IN 96 NORMAL AND 75 SCHIZOPHRENIC *S*s

	Normal		Patient		Probability	% Deficit
	(*M*)	(*SD*)	(*M*)	(*SD*)		
Acceleration latency in seconds						
2° per sec. per sec.	1.3	1.0	3.0	2.8	0.0001	131
5° per sec. per sec.	1.6	1.4	2.1	1.7	0.05	31
12° per sec. per sec.	1.3	1.0	1.7	1.5	0.05	31
No. of acceleration nystagmi						
2° per sec. per sec.	21.6	8.5	11.6	6.8	0.0001	46
5° per sec. per sec.	16.0	6.7	12.2	6.8	0.0002	24
12° per sec. per sec.	19.1	7.0	14.9	7.2	0.0001	22
Deceleration latency in seconds						
2° per sec. per sec.	23.9	3.3	23.0	3.9	0.152	—
5° per sec. per sec.	17.0	2.2	17.2	3.0	0.63	—
12° per sec. per sec.	13.4	1.3	13.8	1.7	0.10	—
No. of deceleration nystagmi						
2° per sec. per sec.	9.0	1.9	6.2	4.0	0.0001	31
5° per sec. per sec.	6.0	3.1	3.9	2.8	0.0001	22
12° per sec. per sec.	4.1	2.3	2.9	1.9	0.0001	34
Rotation rate: nystagmi per second						
2° per sec. per sec.	0.98	0.33	0.60	0.28	0.0001	39
5° per sec. per sec.	1.06	0.36	0.81	0.41	0.0001	23
12° per sec. per sec.	1.47	0.46	1.16	0.50	0.0001	21

TABLE 2

COMPARISON OF POSTROTATION NYSTAGMUS AT 3 STIMULUS INTENSITIES IN 96 NORMAL AND 75 SCHIZOPHRENIC *S*s

	Normal		Patient		Probability	% Deficit
	(*M*)	(*SD*)	(*M*)	(*SD*)		
Number						
2° per sec. per sec.	12.7	7.0	6.1	5.4	0.0001	52
5° per sec. per sec.	14.5	8.2	7.4	5.6	0.0001	49
12° per sec. per sec.	13.8	7.8	8.7	6.2	0.0001	37
Duration in seconds						
2° per sec. per sec.	10.9	4.7	8.0	4.9	0.0001	27
5° per sec. per sec.	11.3	4.5	7.6	4.3	0.0001	33
12° per sec. per sec.	11.0	4.7	8.6	5.5	0.0002	22
Postrotation rate: nystagmi per second						
2° per sec. per sec.	1.16	0.46	0.81	0.46	0.0001	30
5° per sec. per sec.	1.27	0.47	0.97	0.52	0.0001	24
12° per sec. per sec.	1.27	0.47	1.03	0.56	0.0002	19

sity increased, the amount of deficit decreased in the more significant measures, acceleration latency and rotation rate. This inverse relationship between intensity and deficit is seen most clearly in acceleration latency. In the normal controls, the latency of acceleration nystagmus was nearly constant at the mild, moderate, and strong intensities used in the experiment.[4] The latencies of the schizophrenic group, on the other hand, became shorter and approached those of the controls.

Deceleration latency is a special instance of reactivity in which the schizophrenic group showed no significant deficit. Once responding to acceleration stimulation, both groups reversed their reaction to deceleration with equal facility. After reversal had taken place, however, the schizophrenic group at first (during rotation) showed least deficit to the moderate stimulus. In keeping with the general pattern, postrotation deficits were least to the strong stimulus.

Postrotation nystagmus is a phenomenon of receptor adaptation and recovery (Steinhausen, 1933). This after response corresponds to afterimages of rotation which are analogous to afterimages in other sensory modalities. The data in Table 2 amplify in terms of known stimulus intensities the finding of Angyal and Blackman (1941) of postrotation nystagmic deficit in schizophrenia.[5]

DISCUSSION

The direct relationship of nystagmus to the physiology of the semicircular canals has been shown by Steinhausen (1933), Dohlman (1935–36), and Lowenstein and Sand (1940). Their work established angular acceleration as the stimulus. In any head or body movement, eye fixation is maintained by semicircular canal reflexes. In the experimental situation, the fast phase has an adaptive function. Following the eye's slow deviation, the fast phase re-establishes the original fixation. This occurs when the visual field rotates in a constant relationship to the *S*.

An idea of the complexity of the process involved can be gained from a consideration of the events in experimental rotation. Acceleration occurs during the first part and deceleration during the last, when the direction of the slow and fast phases of nystagmus reverses. Deceleration nystagmus perseverates after stimulation (postrotation nystagmus). In a clockwise rotation, the appropriate sensation of rotation to the right is experienced during acceleration. In the brief interval when the stimulus changes from acceleration to deceleration, the rotation seems to stop. There follows the illusion of rotating in the opposite direction. This illusion continues into the postrotation period concomitant with postrotation nystagmus.

---

[4] Differential sensitivity of the right and left labyrinths may account for the slightly longer latency found for the controls at 5° per sec. per sec. than at 2° and 12° per sec. per sec. Rotation was clockwise at 5° per sec. per sec., which ipsilaterally stimulates the right semicircular canals. At 2° and 12° per sec. per sec., the left semicircular canals were stimulated ipsilaterally.

[5] See also Angyal (1941, 1940, 1942).

The results indicate schizophrenic deficits in each stage of this experimental sequence except deceleration latency. They do not indicate where in the neural chain of events (e.g., the semicircular canals) impairment, if any, might lie. The most conservative inferences suggested are those related to the complexity and integration of the neural processes.

1. Of foremost importance is the integrative or adaptive nature of the neural deficit. Considering the proprioceptor and orienting roles the semicircular canals play, the nature of the deficit suggests a biological concomitant of personality disturbances similar to those associated with the body image.

2. The results also emphasize the dimensional nature of deficit in schizophrenia. Nystagmic reactivity for both groups falls into fairly normally distributed measures, and their standard deviations reveal considerable overlap.

3. Adequate neural function was demonstrated in at least one aspect of schizophrenic adaptation. When the stimulus or stress reversed from acceleration (positive) to deceleration (negative), no significant deficit in latency of deceleration nystagmus seemed to occur. Various social and personal aspects of schizophrenia may be subject to a similar principle.

4. With increasing intensity of the stimulus, schizophrenic deficit decreased. This tendency is demonstrated most clearly in latency of nystagmus to onset of rotation. Schizophrenic deficit was less than one-fourth for moderate and strong intensities than for a mild intensity. In its more general implications, this relationship is suggested by stress concepts of Seyle (1956). The greater the intensity of the stress, the closer organisms may approach the uniformity of the general adaptation syndrome.

5. Confirmation of findings of Angyal and Blackman (1940, 1941) of reduced postrotation nystagmus in schizophrenia was experimentally established. Their postrotational mean rates of 1.32 for the patients and 1.68 for the normal controls show a schizophrenic deficit of 21%. Even though the deceleration stimulus produced by the Barany method is estimated at upwards of 1000° per sec. per sec., post rotation deficits are greater at 2° and 5° per sec. per sec. (30% and 24% respectively).

The procedure was especially applicable to mentally ill *S*s with communication difficulties. Minimal cooperation entirely of a nonverbal nature was required of the *S*s. The automatic nature of the response prevented *S*'s rapport from affecting the measurements.

Other advantages were provided by the experimental conditions. None of the *S*s were apparently bothered by aftereffects of rotation. The functional relationship between stimulus intensity and nystagmus could be established because both were precisely quantified. Noteworthy is the comparison of the Barany method to that employed in the experiment reported here. Stimulus levels far below the shock level of that clinical technique were more adequate in identifying the degree of neural integration.

SUMMARY

This experiment demonstrated some aspects of neural integrative adequacies and deficits in schizophrenic function. The nystagmic reflex is a useful measure because of its adaptive nature related to an individual's thresholds, visual-motor orientation, and proprioception. Certain conclusions and implications may be enumerated.

1. Schizophrenics as a group have deficits in latency and amount of nystagmus varying with intensity of stimulation.

2. Adequacy of schizophrenic integration and sensitivity increased as intensity of stimulation increased.

3. No deficit occurred when the stimulus reversed from positive to negative (acceleration to deceleration).

4. A mild stimulus was more effective in demonstrating deficit than stronger stimuli.

REFERENCES

ANGYAL, A. The perceptual basis of somatic delusions in a case of schizophrenia. *Arch. Neurol. Psychiat.*, 1935, **34**, 270–279.

ANGYAL, A. The experience of the body-self in schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1936, **35**, 1029–1053.

ANGYAL, A. Disturbances of activity in a case of schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1937, **38**, 1047–1054.

ANGYAL, A., & BLACKMAN N. Vestibular reactivity in schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1940, **44**, 611–620.

ANGYAL, A., & BLACKMAN, N. Paradoxical reactions in schizophrenia under the influence of alcohol, hyperpnea, and CO₂ inhalation. *Amer. J. Psychiat.*, 1941, **97**, 893–903.

ANGYAL, A., FREEMAN, H., & HOSKINS, R. G. Physiologic aspects of schizophrenic withdrawal. *Arch. Neurol. Psychiat.*, 1940, **44**, 621–626.

ANGYAL, A., & SHERMAN, N. A. Postural reactions to vestibular stimulation in schizophrenic and normal subjects. *Amer. J. Psychiat.*, 1942, **98**, 857–862.

DOHLMAN, G. Toward a method for quantitative measurement of the functional capacity of the vestibular apparatus. *Acta oto-Laryng., Stockh.*, 1935–36, **23**, 50–62.

HOLSOPPLE, J. Q. Space and the non-auditory labyrinth. In C. Murchison (Ed.), *Foundations of experimental psychology.* Worcester: Clark Univer. Press, 1929, Pp. 414–433.

LEACH, W. W. The measurement of nystagmus in normals and schizophrenics. *Dissertation Abstr.*, 1953, **13**, 598–599. Publ. No. 5393, (Univer. Microfilms, Ann Arbor, Mich.)

LOWENSTEIN, O., & SAND, A. The individual and integrated activity of the semicircular canals of the elasmobranch labyrinth. *J. Physiol.*, 1940, **99**, 89–101.

MOWRER, O. H. Concerning the normal function of the vestibular apparatus. *Ann. Otol., Rhinol., & Laryngol.*, 1932, **41**, 412–421.

MOWRER, O. H. The nystagmic response of the pigeon to constant angular acceleration of liminal and supraliminal intensities. *J. comp. Psychol.*, 1935, **19**, 177–193.

SELYE, H. *The stress of life.* New York: McGraw-Hill, 1956.

STEINHAUSEN, W. Über die Beobachtung der Cupula in den Bogengangsampullen des Labyrinths des lebenden Hechts. *Pflueg. arch. ges. Physiol.*, 1933, **232**, 500–512.

TER BRAAK, J. W. G. Über die Empfindlichkeit des Bogengangsapparates für Winkelbeschleunigungen. *Pflueg. arch. ges. Physiol.*, 1936, **238**, 319–326.

(Received November 7, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SELF-ESTEEM AND NEED ACHIEVEMENT AS DETERMINANTS OF SELECTIVE RECALL AND REPETITION[1]

STANLEY COOPERSMITH
Wesleyan University

PERSONALITY theorists ascribe the selectivity of memory to the screening of experiences in terms of their acceptability to the ego. Stimuli that threaten the person's self-esteem are assumed to be less available to consciousness while ones that are acceptable or enhancing are permitted freer expression. Anxiety, threat, stress, and involvement are terms that have been used to describe the psychological state which ensues when there is the possibility of lowered self-esteem. Psychological defenses against anxiety maintain as positive a self-evaluation as is possible, and, as Freud notes, "defenses proceed from the self-respect of the ego" (Freud 1924). Studies of success and failure have suggested (Alper, 1957; Eriksen, 1952, 1954) that where self-esteem is firmly established, failures are more readily recalled and perceived, while insecure or lowered self-esteem results in denial and heightened thresholds for negative stimuli. Similar investigations relating to achievement motivation (Alper, 1957; Atkinson, 1953) strongly suggest that the desire to overcome a failure is a further influence upon the direction of recall, and one that is likely to be expressed in behaviors designed to overcome the affect thereby aroused. The purposes of the present experiment were to investigate the relationship between several indices of self-esteem, achievement motivation, and the recall and repetition of success and failure experiences; and examine the total and positive and negative aspects of n Achievement within the context of self-esteem types previously found to be reliable and consistent (Coopersmith, 1959), and determine the manner in which these aspects are related to these experimental behaviors.

Task interruption has been widely used in experiments concerning selective recall. Zeigarnik (1927) introduced the method with the finding that incompleted tasks were recalled more frequently, and subsequent studies have attempted to determine the variables associated with different recall ratios. Converging lines of evidence now indicate that the recall pattern is a function of the stress introduced into the situation, individual differences in ability to tolerate the stress, and motivations aroused by its introduction. Studies of group trends (Alper, 1946; Glixman, 1949; Rosenzweig & Mason, 1934; Rosenzweig, 1943), although inconsistent, tend to indicate that as threat to self-esteem increases, the recall of incompleted tasks decreases, while the recall of completed tasks remains fairly constant. A number of early investigators remarked upon the personality characteristics of subjects who recalled more failures. They were variously described as proud (Rosenzweig & Mason, 1934), ambitious (Zeigarnik, 1927), and needing success (Cartwright, 1942), and older (Sanford, 1946), and more mature (Rosenzweig & Mason, 1934). Recent studies, more empirical and dynamic in formulation, have investigated personality dimensions broader and more basic in nature. Recall of failure under stressful conditions has been reported for persons high in ego strength (Jourard, 1954; Tamkin, 1957) and achievement motivation (Atkinson, 1953), who show relatively little perceptual defense (Eriksen, 1952) and are more sure of their personal and intellectual worth (Caron & Wallach, 1957; McArthur, 1954). The interaction of situational and personality variables results in four possible recall patterns, two of which, (*a*) a shift from failures to successes under stress, and (*b*) a shift from successes to failures under stress, have been reported. The two other patterns (*c*) recall of successes under all conditions, and (*d*) recall of failures under all conditions, have been suggested (Alper, 1948) but not reported.

In the present experiment subjects differing in subjective and behavioral evaluations of self-esteem, and representing different types of self-esteem (Coopersmith, 1959), were individually presented a series of tasks, half of which were failed. A measure of achievement motivation was then obtained under what were designed to be stressful and arousing conditions. Selective recall and the subject's decisions to repeat either successes or failures were then elicited.

---

[1] The research cited in this paper originated under USPHS Grant M (901) and is currently supported by USPHS Grant M(2696). An earlier version of this paper was presented to the 1959 meeting of the American Psychological Association.

Repetition behavior was obtained on the assumption that it is a more concrete confrontation with previous experience than is recall, and may thus be more anxiety provoking. Furthermore, repetition involves a commitment of future behavior and thus provides an opportunity for testing the hypotheses (Caron & Wallach, 1957; Osgood, 1953) that recall is a cognitive precursor to behaviors that the person has previously found to be constructive. Repetition differs from resumption (Ovsiankina, 1928) in that it controls for differences in point of interruption on failed tasks, and provides an opportunity to redo all tasks in their entirety without distinction between successes and failures.

METHOD

Subjects

The sample consisted of 48 middle middle- and upper middle-class children enrolled in the fifth and sixth grades of an urban elementary school.[2] Four groups, of 12 *S*s each, were chosen from a group of 102 children on the basis of their responses on the Self-Esteem Inventory (SEI), and a teacher's rating of their self-esteem behavior as recorded on the Behavior Rating Form (BRF). The four groups were found to differ significantly on a constellation of variables (achievement, sociometric status, ideal self, and n Achievement) and may be briefly characterized as follows: *High-High* (SEI-BRF)—above average performance commensurate with goals; *High-Low*—low achievers with high goals; *Low-Low*—low achievers who have accepted their inferior performance and position; and *Low-High*—high achievers with extremely high goals but with low self-esteem. It thus becomes possible to analyze patterns of recall and repetition not only in terms of the four types of self-esteem, but also to compare the results of groups high (H-H, H-L) and low (L-L, L-H) in subjective evaluation, and groups rated high (H-H, L-H) and low (H-L, L-L) on their self-esteem behaviors. The measures and methods used to select these *S*s and the elaboration of the self-esteem types have been presented in a previous paper (Coopersmith, 1959).

In deciding which tasks should be employed, the following criteria were established: (*a*) equal attractiveness; (*b*) capable of accomplishment by every child; (*c*) requiring approximately the same amount of time; (*d*) generally familiar; (*e*) interesting; (*f*) success or failure related to the *S*'s performance; and (*g*) relative unlikelihood that *S*s would interpret their accomplishment in terms of partial success and failure. In general, it was found that tasks requiring simple motor coordination were most effective.

In order to control for recall or reptition due to the nature or attractiveness of the task, a success and failure variation of each task was employed. Only those tasks rated approximately equal in attractiveness were used. Although this use of alternates of the same task would tend to increase the likelihood that the first would act as a mnemonic to the second, the tasks were sufficiently different and presented in such an order and manner that their relatedness was not clearly apparent. The six pairs of tasks consisted of: (*a*) Mosaics—*S* was asked to replicate a design of a boy and a flowerpot using plastic heptagonal blocks; (*b*) Cut-outs—*S* was shown a paper cut-out of an animal after which the *E* disassembled it and asked the *S* to reproduce it; (*c*) Scrambled sentences—*S* was asked to arrange the words presented to him on cards into a meaningful sentence; (*d*) Kohs blocks—*S* was asked to reproduce a design of a house and a series of stripes; (*e*) Jig-saw puzzle—*S* was shown a completed puzzle after which the *E* disassembled it and asked the *S* to reassemble it; and (*f*) Snap-block construction—*S* was asked to reproduce a design of a horse and a robot using varied shaped blocks that snapped together.

Procedure

Task administration. The set of 12 tasks was administered individually to each *S* under conditions designed to threaten self-esteem. *E* had previously announced to the class that they were to participate in a study designed to test the attitudes and achievement of children. The selected *S* was taken to a private office in the school building, seated before a desk, at the side of which sat *E*, stopwatch in hand. *S* was informed that he would be permitted two minutes to complete each task and that *E* would time him and record the nature of his performance. The tasks were out of *S*'s sight until *E* presented them individually in the sequence S F S S F F S F F S S F. They were rotated so that success tasks occurred an equal number of times in each success position and failure tasks an equal number of times in each failure position. Success and failure were controlled by timing, with the difference between success and failure variations of each task generally less than 30 sec. Analysis indicated no effect due to either serial position or point of interruption. At the conclusion of each task *E* announced that *S* had either succeeded or failed upon the task; he repeated this before proceeding with the next task.

Measurement of n *Achievement.* After *E* had placed the last task out of view, he asked *S* to complete four stories that had previously been found (McClelland,

---

[2] The author wishes to thank the superintendent, principals, faculty, and students of the Ithaca Public School System for their cooperation with this study.

Atkinson, Clark, & Lowell, 1952) to elicit n Achievement in children. This measure was included to determine motivation under arousal conditions. It provided a "filler" between performance and recall, and made it possible to evaluate the relationship between achievement motivation and recall as previously demonstrated (Atkinson, 1953).

The story completions dealt with: a boy (or girl) at school; brothers and sisters playing; two men (or women) near a machine; and a boy (or girl) seated at a desk. As the child presented the stories orally, they were recorded by *E*. The average duration of this period was approximately 15 min. It had previously been determined (McClelland et al., 1952) that stories were more successful than pictures in eliciting achievement motivation from children.

The scoring of the stories followed the established criteria (McClelland et al., 1952), each story being scored separately for the presence or absence of each of the 13 achievement related categories. They were further grouped into the *positive* (I+, Ga+, G+, Achievement Thema) and *negative* aspects (I−, Ga−, G−, B) subsumed under the total n Achievement score, aspects which are apparently indicative (McClelland et al., 1952) of hope of success and fear of failure respectively. A product-moment correlation of .85 was obtained between the total n Achievement scores of two independent judges.

Obtaining recall and repetition. After *S* had completed the last of his stories, *E* asked him to report which of the tasks he remembered most clearly of all upon which he had previously performed, then as many others as he could recall. Pretests had established that a large percentage of the children could recall all the tasks and it was therefore decided to include a measure of "clearest" recall, with instructions indicating the acceptability of any recall preference.

E then led the child behind a curtain to where the tasks, all completely assembled, were located. He indicated a pair of tasks, offered *S* an opportunity to repeat either the success or failure variation, and after the decision had been made allowed him to redo the task. Decisions were obtained for the other pairs of tasks, although *S* did not actually undertake the task again in the remaining five cases. Before leaving he was asked to keep the entire procedure in confidence. Insofar as is known, no breach of confidence occurred.

TABLE 1
MEAN NUMBER OF SUCCESSES RECALLED BY DIFFERENT TYPES OF SELF-ESTEEM

Groups[a]	Initial	First Three[b]	Total[b]	Fail/Succ. Ratio
Low-Low	0.8	2.2	4.7	.84
Low-High	0.5	2.2	5.1	.89
High-Low	0.5	2.0	4.8	.91
High-High	0.8	2.3	5.3	.83
P[c]	<.30	>.30	>.30	—

[a] The terms refer to upper vs. lower quartile status on the Self-Esteem Inventory and the Self-Esteem Behavior Rating Form, respectively.

[b] The individual scores for the first three, and total recalls were transformed into percentage success form. The recall percentages were then transformed into angles by the arc sine percentage transformation.

[c] *P* is based on χ^2 distribution for initial recall, and on the F distribution for the first three and total recalls.

TABLE 2
FREQUENCY OF RECALL AND REPETITION PATTERNS FOR SELF-ESTEEM TYPES

Group	Recall			Repetition		
	More Fail.	Equal	More Succ.	More Fail.	Equal	More Succ.
Low-Low	2	4	6	1	2	9
Low-High	1	5	6	9	2	1
High-Low	3	3	6	8	2	2
High-High	0	5	7	1	4	7

RESULTS

Self-Esteem

In Table 1 the recall behaviors of the different types of self-esteem are compared. In order to test for a possible sequential effect, initial recall, first three recalls, and total recall of successes were noted. The most striking feature of these results is the lack of any significant differences between the self-esteem groups. The Low-High and High-Low groups showed a tendency to recall more failure on the first recall but this difference is well within chance expectancy. On the first three recalls all groups recalled at least twice as many successes as failures, while the ratio of failure to success of less than one for total recall corroborates the general tendency for the recall of success tasks. The possibility that *S*s in each group would differ in the frequency with which they recalled more successes, more failures, or an equal number of each is refuted by Table 2. Chi square for the distribution of *S*s falling into each category was clearly not significant.

Comparison of the groups high (H-L, H-H) and low (L-H, L-L) in subjective evaluation reveals no differences on any of the recall measures, nor does comparison of the groups high (H-H, L-H) and low (L-L, H-L) on ratings of self-esteem behavior. The present findings, then, do not support the view that selective recall is related to subjective or behavioral indices of self-esteem, nor do children representing different types of self-esteem differ in their recall patterns. These results are consistent with previous studies

TABLE 3

CORRELATIONS OF MEASURES RELATED TO RECALL AND REPETITION

Variable	No. of Successes Recalled	No. of Successes Repeated
Number of successes recalled	—	$-.26$
Self-Evaluation (SEI)	$-.21$	$-.25$
Behavior Evaluation (BRF)	$-.03$	.13
Objective Achievement (Iowa)	$-.08$	.23
D(SB)[a]	$-.08$	$-.27$
Manifest Anxiety	.08	$-.11$
Ideal Self	.02	$-.17$
Need Achievement	.12	$-.38$
D(SI) adjusted[b]	.16	.07

Note.—Value of r for $N = 48$, .05 level = .28, .01 level = .37.

[a] Difference between Self and Behavioral Evaluations.

[b] Difference between Self and Ideal Self, adjusted for "ceiling" effect.

(Caron & Wallach, 1957; Glixman, 1949; Smock, 1957) in the finding that there is a greater recall of successes than failures under stressful conditions.

For task repetition, the dependent variable is the total number of success tasks repeated. Table 2 indicates that *S*s in the Low-High and High-Low groups tended to repeat failures while Low-Lows and High-Highs repeated successes. Thus, groups showing agreement between SEI and BRF scores repeated twice as many successes as failures, while cases of disagreement repeated failures twice as often as successes. The comparison of groups high and low in self-evaluation, and the groups high and low on their behavioral ratings show no differences in repetition.

In the course of preparing the profiles of the self-esteem types (Coopersmith, 1959) information on several personality and motivational variables had been obtained. A correlational analysis was performed to determine the nature and extent of the relationships between these measures and the recall and repetition of successes. The results of this analysis are presented in Table 3. The analysis includes data based on three scores derived from differences between measures. The first of these is the D(SB) score, based on the difference between subjective and behavioral evaluations. Since the groups with high positive (H-L) and high negative (L-H) D(SB) scores behaved in identical fashion in the repetition experiment, the direction of the discrepancy will not be considered in this analysis. The D(SI) and the D(SI) adjusted scores are based upon the difference between subjective and ideal self-statements, with the latter score adjusted to correct for the ceiling imposed on the D(SI) score by persons high in their self evaluation. Since the correlation between the two D(SI) scores was significant ($r = .64$, $p < .01$) only the adjusted score is included.

None of the variables in Table 3 appear to be significantly related to the number of successes recalled. The only correlation approaching significance was between the number of successes recalled and number of successes repeated ($r = -.26$, $p < .08$). Of the four other variables related to the repetition of successes, achievement motivation was clearly significant ($p < .01$), while the D(SB) score ($p < .06$), self evaluation ($p < .09$), and objective achievement ($p < .10$) approached accepted significance levels. Partial correlations indicate that n Achievement and the D(SB) score, the two best predictors, are each independently related to repetition behavior. The beta coefficients of the multiple correlation of repetition with the n Achievement, D(SB), and SEI scores combined ($r = .49$, $p < .01$), signify that these variables contribute respectively 45%, 34%, and 21% of the relationship. The coefficient of multiple determination ($R^2 = .2349$) indicates that these variables combined can account for 23% of the variation in repetition.

Achievement Motivation

The finding that the two self-esteem types (High-Low and Low-High) that are high in achievement motivation but significantly different in other characteristics, both repeat failures, raises the question of whether their actions are based upon similar underlying motives or upon different motives which produce the same effect. A number of studies (Alper, 1957; Atkinson, 1953; McClelland et al., 1952) indicate that the total n Achievement score includes a positive aspect indicative of an approach orientation and a motivation to overcome obstacles, and a negative aspect oriented toward avoiding failure. On the basis of the curvilinear relationships found in studies of perception (McClelland et al., 1952) and recall (Atkinson, 1953), it has been proposed (McClelland et al., 1952) that negative responses occur with greater relative frequency in the middle and lower portions of the distribution of total n Achievement scores.

Examination of Table 4 indicates that there are significant differences between the groups in positive and negative subcategories as well as in the general categories. While all groups have more responses in the positive than in the negative categories, the Negative/Positive ratio indicates that the High-Highs have the highest proportion of positive responses (and are thus presumably approach-oriented), the Low-Lows the highest proportion of negative responses (and are thus presumably avoidant-oriented), while the two groups showing discrepancies between self-ratings and behavior have more of a balance between positive and negative categories, a situation which would presumably lead to conflict. These findings are consistent with the personality characteristics and dynamics of the self-esteem types (Coopersmith, 1959) and support the belief that they represent distinct character structures.

Inspection of Table 4 further reveals that of the two groups which are low in achievement motivation the Low-Low self-esteem group shows the greatest percentage of negative responses, whereas the High-Highs show the greatest percentage of positive responses. The findings indicate that the positive and negative categories may vary independently of each other and of the total n Achievement score.

TABLE 4

TOTAL NUMBER OF RESPONSES OF THE SELF-ESTEEM GROUPS IN GENERAL CATEGORIES AND POSITIVE AND NEGATIVE SUBSCORES

Group	General Categories			Total n Ach and Subscores			Ratio, Neg./Pos.
	AI[a]	TI[b]	UI[c]	Total	Negative	Positive	
Low-Low	22	15	11	48	16	19	.84
Low-High	33	5	10	105	26	47	.55
High-Low	23	12	13	72	21	34	.62
High-High	19	12	17	49	8	32	.25
	*	*	—	**	*	*	**

* $p < .05$.

** $p < .01$.

[a] Achievement Imagery—Indicated by long term involvement, unique accomplishment, or competition with a standard of excellence.

[b] Task Imagery—Some reference to achievement but not explicit in terms of AI criteria.

[c] Unrelated Imagery—No references to achievement.

TABLE 5

POSITIVE AND NEGATIVE SUBCATEGORIES OF ACHIEVEMENT MOTIVATION BY SELF-ESTEEM GROUPS

Positive	I+	Ga+	G+	Ach Th	Tot. Pos.	Pos-Ach Th
Low-Low	9	3	7	0	19	19
Low-High	14	8	14	11	47	36
High-Low	12	6	11	5	34	29
High-High	10	3	11	8	32	24
	—	—	—	*	*	—

Negative	I−	Ga−	G−	Bw	Bp	Tot. B	Tot. Ng	Neg.-Bw
Low-Low	0	3	11	1	1	2	16	14
Low-High	—	6	19	0	1	1	26	26
High-Low	2	4	7	6	2	8	21	15
High-High	0	1	6	1	0	1	8	7
	—	—	—	*	—	*	*	*

* $p < .05$.

** $p < .01$.

Note.—Key to subcategories:

Positive
- I+ Successful instrumental activity
- Ga+ Anticipation of success
- G+ Positive affect related to achievement
- Ach Th Achievement imagery central theme

Negative
- I− Unsuccessful instrumental activity
- Ga− Anticipation of failure
- G− Negative affect related to achievement
- Bw Environmental obstacle
- Bp Personal obstacle

With respect to responses within the subcategories, Table 5 indicates that the major difference for the positive subcategory is in Achievement Thema responses while that of the negative aspect is in Bw responses representing impersonal obstacles to achievement. The difference in the negative category remains significant when the Bw responses are removed, whereas the difference in the positive category no longer holds with the removal of Achievement Thema responses.

In order to take account of the differences between the groups in Achievement related stories, the total positive and negative categories as well as the specific responses for each *S* were divided by the number of his AI stories. This analysis indicated significant differences for only the total negative category and the Achievement Thema and Bw responses.

Correlational analysis of the achievement motivation subcategories with recall and repetition revealed a significant negative relationship between total positive responses and the number of successes repeated ($r = -.36$, $p<.01$). The correlation for the total positive alone ($r = -.36$) is similar to that attained for the total n Achievement score ($r = -.38$). None of the other correlations between categories and responses and the experimental behaviors, whether by type or for the total group, achieved accepted levels of significance, thus raising the possibility that where differences were obtained they could be accounted for by chance. A further limit on interpreting the relationship between n Achievement and repetition is set by the recognition that the groups which were high in their D(SB) scores were also high in n Achievement, thus raising the possibility that the findings could be due to the effects of either variable, of neither, or of yet another variable as yet unidentified. Interpretation must therefore await experimental evidence indicating the significance of the self-behavioral discrepancy, and the independent effects of the n Achievement and D(SB) variables.

DISCUSSION

The results of the present study clearly do not support the hypothesis that either achievement motivation or self-esteem are related to recall preference. The general result indicating a greater recall of successes under stress is in agreement with the results of previous studies (Glixman, 1949; Rosenzweig, 1943; Smock, 1957), despite a bias introduced in the sample selection. Of those cases where a greater recall of failures under stress has been reported, only one (Atkinson, 1953) indicates an absolute majority of failures, while the others refer to either the recall pattern of particular individuals (Alper, 1946; Atkinson, 1953; Caron & Wallach, 1957; Eriksen, 1952) or to an increase relative to some previous finding (Alper, 1948; Glixman, 1949; Rosenzweig, 1943; Smock, 1957). Thus, while the suggested relationships of this study are not borne out, the group results support the general finding of a recall bias favoring success.

It is this very absence of relationships between the battery of variables and recall, as well as the small proportion of individuals in the sample who recall more failures, which raises serious yet intriguing questions. Turning first to situational and task variables for an explanation, it may be argued that *S*s in the present experiment may have been placed under greater stress than is generally the case, and since one of the effects of stress is to decrease the number of failures recalled (Glixman 1949; Rosenzweig, 1943; Smock, 1957) superordinate stress should presumably accentuate this trend. While this may have been the case, *E*'s observations and the comments by the *S*s did not indicate an arousal greater than that reported by others investigators. In addition, recall experiments differ on such variables as the number of tasks used (4–30), the method of administration (group or individual), the type of task (scrambled sentences, drawing, puzzles, problems, etc.), the length of tasks, the length of time between performance and recall, and whether or not a "filler" activity is introduced.

A further reason that may be advanced is that the use of alternates of the same task may have acted as a mnemonic and thereby confounded the recall pattern by increasing it *in toto*. Here the matter would seem to be one of strategy since it is clearly necessary to control for the nature and attractiveness of the task, and while each of the tasks may well act as a mnemonic for the other, procedures which do not control this variable are subject to the contrary rejoinder. While these differences make comparisons difficult, it should be noted that the procedures employed in the present study are essentially similar to those employed in studies reporting a much larger percentage of *S*s recalling more failures than successes (Alper, 1946; 1957; Caron & Wallach, 1957; Glixman, 1949; Jourard, 1954; Tamkin, 1957).

The most likely explanation for the present findings appears to be that the *S*s were younger than those generally employed in recall studies. It is clearly not the passage of time but rather the quantity of experiences as well as their quality (success and failure) which is presumably related to recall preference. The two characteristics linked to maturity that appear to be related to recall are: (*a*) the development of a tolerance for the anxiety aroused by failure and a perspective upon its significance, an ability subsumed under such terms as ego strength and self-esteem and (*b*) an attitude towards failure as frustrating and insulting which results in behavior designed to change it to success, a measure of which is apparently obtained in the achievement motive. Persons high in both these characteristics should recall failures under stress and nonstress, while those low in both these characteristics should recall successes under both conditions. Persons with intermediate amounts of the two, in varying combinations, may be expected to recall either success or failure, depending on the extent to which the situation is perceived as failure and the extent to which n Achievement is aroused by situational cues.

A good deal of the controversy as to whether the recall of failures is indicative of a weak (Alper, 1946; Rosenzweig & Mason, 1934) or strong ego (Sanford, 1946) appears due to a failure to distinguish between the concepts of *appropriateness* and *strength.* Appropriateness refers to a response suitable to a particular situation and may be acquired through relatively superficial social learning, while psychological strength refers to the ability to tolerate adversity and less need to distort and restrict reality, thus increasing the probability that adversity will be overcome. The former is based upon familiarity with a given environment and perceptual sensitivity to its cues, while the latter is achieved by states of high motivation, a wide range of experiences, and the confidence gained in overcoming adversity. The two clearly are not synonymous and it is plainly not possible to reason from one to the other.

Failure to distinguish between the appropriateness and strength in the case of recall leads to the position that the distortion and restriction of reality which results from protecting self-esteem is an indication of greater strength than accurate perception and a corrective response. It also leads to the conclusion that children have stronger egos than do adults, a conclusion clearly unwarranted by current theory or evidence. A more meaningful interpretation would appear to be that adults have learned to act more appropriately than children. Accordingly, they may be expected to report responses publicly in accord with group expectations, ones that do not necessarily correspond to the total or most vivid recall experienced.

In evaluating the hypothesis (Caron & Wallach, 1957; Osgood, 1953) that recall is a rehearsal of future behavior with a view towards commitment, it is necessary to distinguish between those cases of repetition in which the *S* is confronted with the various tasks and asked to choose between them, and those in which the repetition choices must come from the tasks which he has recalled. The former case allows a wider psychological world to operate than does the latter, which may provide alternates for repeating none of the variables, only one, or a choice among several variations. It is obvious that if the person is motivated to overcome a failure experience but cannot recall it, repetition is an impossibility unless he is confronted by it. Since the procedures for aided and unaided repetition may produce different results and the present test of the rehearsal hypothesis is limited to the condition of aided repetition, it may be suggested, in view of the relatively low correlation ($r = .26$) between recall and repetition, that if recall is a rehearsal, it is not one that frequently results in performance. Since the results indicate that more successes are recalled than failures, unaided repetition, i.e., that based solely on the recall of the *S*, would appear to be more likely to result in a greater repetition of successes than would aided repetition.

It should be borne in mind that aided repetition is more anxiety provoking since it forces upon the individual the recognition that he has failed, something which can be denied when the defenses are allowed full sway. In terms of failure repetition, which would appear under all circumstances to cope more effectively with reality, the Low-High and the High-Low groups apparently are more capable and willing to undergo anxiety in order to attain success than are the High-High and Low-Low *S*s who continue to repeat successes even when confronted with failures.

SUMMARY

This study investigates the relationship between self-esteem and the recall and repetition of success and failure experiences. Four groups, of 12 *S*s each, were selected on the basis of their self-evaluative responses and an evaluation of their self-esteem behaviors. The combinations of these two variables, at their

extremes, yielded significantly different patterns on such variables as achievement, ideal self, and sociometric status, and apparently represent distinct types of self-esteem.

Each *S* was given a series of six pairs of tasks, in an achievement oriented situation. The *S* failed one of each pair of tasks and succeeded on the other. The tasks were followed by four story completions which had been found to elicit n Achievement. Recall was then obtained, after which the *S* was offered an opportunity to repeat either the success or failure variation of each task.

The data were analyzed both in terms of the four groups differing in self-esteem and by a correlational analysis to determine which variables were related to recall and repetition. All groups recalled more successes, and none of the independent variables was significantly related to recall. Groups in which self-rated self-esteem was discrepant from behavior ratings tended to repeat failures; those in which the two measures agreed tended to repeat successes. An index of agreement between self-rated self-esteem and behavioral self-esteem and n Achievement were the variables most significantly related to repetition. Analysis of the need achievement scores indicated differences between the self-esteem types in positive, negative, and general categories as well as subcategory responses, although none of these were related to the experimental behaviors.

It is proposed that two factors are necessary for the recall and repetition of failure. These are the ability to tolerate failure and the motivation and striving to overcome its effects.

REFERENCES

Alper, Thelma G. Memory for completed and incompleted tasks as a function of personality: An analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403–420.

Alper, Thelma G. Task orientation and ego orientation as factors in reminiscence, *J. exp. Psychol.*, 1948, **38**, 224–238.

Alper, Thelma G. Predicting the direction of selective recall: Its relation to ego strength and n Achievement. *J. abnorm. soc. Psychol.*, 1957, **55**, 149–165.

Atkinson, J. Achievement motivation and the recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

Caron, A. J., & Wallach, M. A. Recall of interrupted tasks under stress: A phenomenon of memory or of learning? *J. abnorm. soc. Psychol.*, 1957, **55**, 372–381.

Cartwright, D. The effect of interruption, completion, and failure upon the attractiveness of activities. *J. exp. Psychol.*, 1942, **31**, 1–16.

Coopersmith, S. A method for determining types of self-esteem. *J. abnorm. soc. Psychol.*, 1959, **59**, 87–95.

Eriksen, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Physchol.*, 1952, **47**, 230–235.

Eriksen, C. W. Psychological defenses and "ego strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.

Freud, S. On narcissism: An introduction. In *Collected papers*. Vol. 4. London: Hogarth, 1924.

Glixman, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.

Jourard, S. M. Ego strength and the recall of tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 51–58.

McArthur, C. Personalities of public and private school boys. *Harvard educ. Rev.*, 1954, **24**, 256–262.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1952.

Osgood, C. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

Ovsiankina, Maria. Investigations on the psychology of action and affection: VI. The resumption of interrupted acts. *Psychol. Forsch.*, 1928, **11**, 302–379.

Rosenzweig, S. & Mason, G. An experimental study of memory in relation to the theory of repression. *Brit. J. Psychol.*, 1934, **24**, 247–265.

Rosenzweig, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustrations. *J. exp. Psychol.*, 1943, **32**, 64–74.

Sanford, R. N. Age as a factor in the recall of interrupted tasks. *Psychol. Rev.*, 1946, **53**, 234–240.

Smock, C. D. Recall of interrupted and non-interrupted tasks as a function of experimentally induced anxiety and motivational relevance of the task stimuli. *J. Pers.*, 1957, **25**, 589–599.

Tamkin, A. S. Selective recall in schizophrenia and its relation to ego strength. *J. abnorm. soc. Psychol.*, 1957, **55**, 345–349.

Zeigarnik, B. On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Harcourt Brace, 1938, Pp. 300–314.

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

COOPERATIVE BEHAVIOR IN DYADS AS A FUNCTION OF REINFORCEMENT PARAMETERS[1]

SEYMOUR ROSENBERG

Bell Telephone Laboratories, Murray Hill, New Jersey

STUDIES of cooperation in the laboratory have been stimulated, in part, by the large number of life situations in which cooperative action is observed both to occur and to break down. Biologists and ethologists (Lindzey, 1954, pp. 540–541; Tinbergen, 1953) have isolated and analyzed a number of examples of cooperative behavior among the lower organisms. These relationships, termed symbiotic or commensal, have frequently been noted in dyads for feeding or protective purposes and are often composed of members from different species. Elaborate systems of cooperative behavior as well as breakdowns in these systems also pervade human activity (Lindzey, 1954, p. 754; Mintz, 1951). Many cooperative situations are difficult to analyze with currently available techniques and theory. The abstraction and isolation of variables in the laboratory should enhance our understanding of these complex phenomena. The purpose of this study is to describe a number of reinforcement parameters in terms appropriate to dyadic cooperation and to study these parameters experimentally. The choice of reinforcement parameters for the present study is based upon the general conception that the reinforcing events are critical components in the development and maintenance of cooperative action.

Even a cursory review of the literature on cooperation reveals a number of different usages of the concept (Deutsch, 1949). To avoid ambiguity, a definition of a cooperative situation is adopted for this study that does not attempt to include all possible uses of the term, but does share an area of meaning with other definitions in the literature and is intended to be applicable to a variety of experimental conditions. The definition is formulated for dyads but extensions to larger groups are easily made. A cooperative situation has all of the following four characteristics: (*a*) There exist at least two outcomes for each individual following a response. (*b*) At least one of these outcomes is rewarding (i.e., it provides positive reinforcement and/or terminates noxious stimulation), and at least one outcome is not rewarding (or less rewarding). (*c*) Reward delivered to one person must not preclude its availability for the other. In this study a reward was available to both persons simultaneously during each trial. A cooperative situation may be contrived, however, in which reward is deliverable to only one or the other person in any one trial or time period. If the study consists of only one trial, the latter situation would be termed competitive rather than cooperative. (*d*) The outcomes to both persons are, in part or entirely, a function of the response of the other person. That is, the outcome to Person X must depend to some extent upon the behavior of Person Y *and* the outcome to Person Y must depend to some extent upon the behavior of Person X. If the outcome to only one is dependent on the behavior of the other, it is only a "partially cooperative situation." It resembles a commensal rather than a symbiotic relationship. There may, however, be additional asocial determinants for the outcomes in a cooperative situation, e.g., a random component.

Cognitive concepts are not included in this definition of a cooperative situation. "Awareness," defined either in terms of experimenter operations (e.g., instructions to the *S* that it is a "cooperative situation") or in terms of the *S*'s behavior, (e.g., questionnaire to the *S* to determine his verbalizations about the cooperative aspects of the situation) is not a necessary condition for the specification of a cooperative situation. Nor is it always a necessary condition for the development and maintenance of cooperative action (Hall, 1957;

---

[1] This research was performed by the author at Lackland Air Force Base, Texas. This research was done under ARDC Project No. 7737, Task No. 27006, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Gov.

Rosenberg & Hall, 1958; Sidowski, Wyckoff, & Tabory, 1956; Sidowski, 1957). That is, the consistent occurrence of response combinations that produce optimally rewarding outcomes for both persons does not require awareness in either sense of the term. The exclusion of awareness as a necessary condition for the definition of a cooperative situation, however, implies nothing about its relevance to the behaviors that take place in some cooperative situations.

The experimental situation adopted in this study has been used in a number of previous studies (Hall, 1957; Rosenberg, 1959; Rosenberg & Hall, 1958). While these studies were concerned primarily with the development and maintenance of precise motor activity in a team setting, the task and procedures were also found to be quite applicable to the problems discussed in this paper. Each person was given a series of trials in which he performed the simple motor response of turning a small concealed knob. It is a trial-and-error response in that the only discriminative stimulus for each trial was a signal to respond. The reinforcing stimulus was delivered to each *S* after each trial. Selected parameters of these reinforcing stimuli were considered important to the development and stability of cooperative action. These parameters were varied experimentally in a series of three experiments. The first experiment deals exclusively with cooperative situations. The other two experiments contain some conditions that are only partially cooperative. All conditions and every trial, however, involved a combination of outcomes that was positively reinforcing to both *S*s. The degree to which *S*s were able to learn to procure these simultaneous reinforcements, i.e., develop cooperative action, was studied.

The task used in these studies restricts and simplifies the interpersonal actions that occur to a much greater degree than has been customary in social psychology. Special emphasis has been placed on controlling the way in which one person's responses may affect another's stimulation. The control of these linkages appears to be central to extensions of S-R descriptions to interpersonal events, at least when viewed from our present state of understanding. Similar attempts to extend S-R theory have been expressed in the experimental studies of Sidowski et al. (1956) and Sidowski (1957) in their use of a "minimal social situation" and in earlier discussions by Miller and Dollard (1941), and by Logan (1955).

Method

Apparatus

Two identical adjacent booths were provided, one for each *S*. The walls of the booth were sufficiently high so that *S*s could not see each other. One wall of each booth consisted of a panel from which a small cylindrical knob projected about ½ in. Each *S* was provided with his own panel and knob. The knob was mounted about 3 ft. above floor level and was concealed by a wooden shield. The *S* could turn the knob by reaching under the shield but had no visual access to the knob. The knob was homogeneously knurled on its entire surface to avoid giving tactual cues of the amount it was turned. The *E*, seated on the other side of the panel, could read from a dial the amount each *S* turned the knob on each trial.

On the panel in front of each *S*, about a foot above the shielded knob, two or three 6-w. lights (depending upon the experimental condition) were mounted about 1.5 in. apart and labeled so as to provide a reinforcing stimulus after each trial.

The knob apparatus is essentially the same as that described in a previous publication (Rosenberg & Hall, 1958). In this study, however, reinforcing stimulus values were two- or three-valued rather than scalar quantities.

Subjects

*S*s were male Air Force basic trainees whose modal age was about 17 years. Trainees were assigned as a unit for use as *S*s in psychological research but were excluded form the experiment if their Armed Forces Qualification Test (AFQT) score was below the 50th percentile rank, which is roughly equivalent to a Stanford Binet IQ of 100 (Gragg & Douglass, 1956). *S*s were used in pairs, randomly assembled. Twenty pairs were used for Experiment 1, 60 pairs in Experiment 2, and 36 pairs in Experiment 3.

General Procedure[2]

*S*s in all experiments and in all experimental conditions were given the same instructions followed by 50 trials. Instructions were tape-recorded and heard by *S*s through earphones. An *E* pointed to and operated appropriate parts of the panel when they were mentioned in the instructions. Each *S* was told:

> The Air Force needs to know more about how well people can do certain tasks. The experiment in which you are to take part is a study of just one

---

[2] The author wishes to acknowledge the able assistance of Richard W. Hall and Stuart Dobson in the conduct of the experiments. Their dedicated efforts made it possible to complete this series of experiments.

such task. It is a simple task of turning a knob the right amount. Listen carefully to these instructions and the experimenter will point out things as they are mentioned. On the panel in front of you is a shield, and underneath the shield where you cannot see it, is a small knob. You may look under the shield now to locate the knob, but you are expected to work without seeing the knob. *Your job is to learn to turn the knob just the right amount. You will have a number of chances to do this. The green signal light will tell you when to start and stop. When the green light comes on, you begin to make your setting. As soon as you have made the setting you think is correct, take your hand away from the knob. When the green light goes off you must take your hand away whether you are finished or not. You must turn your knob each time the green light comes on. How much you turn it is what you have to learn. You will find that you have plenty of time without rushing. A few seconds after you have made your setting, you will be shown how well you did by the signal lights mounted just above the shield. Always turn your knob to the right as indicated by the sign tacked just over the shield. After each trial your knob is set back by the experimenter so that each time you make a setting you are starting from zero.* [long pause] Because it is important that you understand these instructions, I will repeat the most important parts. [Italicized portion repeated here.] Now we are ready to start. Do not touch your knob until the green signal light comes on. Now remove your earphones and watch for the green light to come on.

One complete trial consisted of three automatically timed intervals: (*a*) a response period of 8 sec. indicated to the *S* by a green signal light during which the knob could be turned; (*b*) a delay of 10–12 sec. during which *E* recorded the response magnitude of each *S* in the dyad, set these values into a computer, and read from the computer the values of reinforcement to deliver to each *S*; (*c*) the presentation of reinforcement by turning on the appropriate lights of *S*'s panels for 2 sec. After 2–3 sec. delay a new trial was started.

The presence of both *S*s in the same room did not appear to suggest to any of the *S*s that a behavioral link might exist between their tasks. A sizable number of *S*s made spontaneous comments about the task and "how it works" at the end of the session but none commented about the possibility of an interpersonal linkage. The nature and presentation of the instructions, the use of separate, completely enclosed booths for each *S*, the absence of an apparent mechanical linkage between the apparatus in the two booths all seemed to create the set for each *S* that he was being studied as an individual, albeit simultaneously with another person.

EXPERIMENT 1

Experimental Conditions

Each member of the dyad was assigned to one of five social reinforcement conditions. These five conditions differed from each other in that a different functional relationship was used to determine the S^r (reinforcement stimulus) to an *S* from the response magnitudes of both S_x and S_y.

To understand the derivation of the five reinforcement conditions, it is first necessary to define a variable Z. The five conditions were generated by assigning one of five values to the parameter, a_x *or* a_y, in the following equations

$$Z_x = a_x R_x + (1 - a_x) R_y$$
$$Z_y = (1 - a_y) R_x + a_y R_y$$

where $1 \geq a \geq 0$, and where R_x and R_y are the numbers of turns by each *S* on a particular trial. Note that the Z function is a simple weighted average of the two response magnitudes obtained on each trial and that the value of *a* is the proportion of variability in the Z values of an *S* that is attributable to his own response. If a_x were 1.00, for example, the situation would be reduced to individual trial-and-error learning for S_x. He would determine all of the variability of the Z values. All other values of *a* would be considered "social."

The numerical values of Z_x and Z_y were not the actual reinforcement stimuli delivered to S_x and S_y, respectively. The numerical value of Z on a trial determined which one of the following three S^r events were presented to an *S* on that particular trial:

1. Poor—Too Little	when $Z < 3.6$ turns
2. Good	when $Z = 4.0 \pm \leq .4$ turns
3. Poor—Too Much	when $Z > 4.4$ turns

This is essentially a coarse grouping of Z values, in which weighted average values below or above a certain range are Poor and with further specification as to direction. The same range of Z, $4.0 \pm \leq .4$ was used to transform both Z_x and Z_y.

When the parameters of both functions, Z_x and Z_y, have been specified numerically, and the relationship between Z values and S^r events stated, an "interpersonal reward system" is said to have been constructed.[3] In Experiment 1, five interpersonal reward systems were constructed by using the five *a* values: .55, .525, .50, .475, and .45, and im-

---

[3] Note that the term "reinforcement condition" refers to the operations by which an S^r to a single *S* is determined and that "interpersonal reward system" refers to the *combination* of "reinforcement conditions" for two or more persons. This terminology is used consistently throughout the paper.

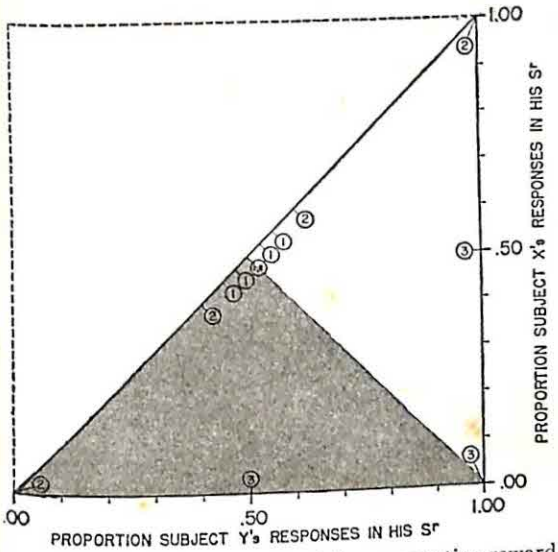


FIG. 1. Graphic summary of the cooperative reward systems used in the three experiments. The circled numbers represent the combinations of conditions for S_x and S_y in each experiment. The circled Digit 1 represents the combinations in Experiment 1, the Digit 2 in Experiment 2, and the Digit 3 in Experiment 3. The other half of the square would be a repetition of the same combinations with the conditions reversed for S_x and S_y.

posing the restriction that $a_x = a_y$ for each reward system. Thus, Experiment 1 was limited to reward systems in which each member of the dyad received the same *proportion* of *his own* response in his reinforcement. That is, when a_x was .525, for example, a_y was also .525. This should be carefully distinguished from a system in which both members receive the same reinforcement *values*, (e.g., $a_x =$ .525, $a_y =$ .475). Figure 1 is a graphic summary of the combinations of a_x and a_y chosen in Experiments 1, 2, and 3. The five combinations of a_x and a_y used in Experiment 1 are indicated by the figure 1 in a circle. The rationale for choosing these particular functions was derived from previous study. Previous data (Rosenberg, 1959) suggested that these five points represent the critical range of the continuum from those systems in which procurement and maintenance of reward develop for both *S*s to those systems in which continuous reward failed to develop for at least one of the two *S*s.

It should be noted that all five reward systems in Experiment 1 satisfy the definition of a cooperative situation on every trial. That is, there exist at least two outcomes for each *S* following a response, at least one is rewarding (Good) and at least one is not (Poor), reward to one *S* does not preclude its availability for the other, and, finally, the *a* values chosen determine that the outcomes to both *S*s are, in part or entirely, a function of the response of the other *S*. In fact, all combinations of a_x and a_y in Fig. 1, except where a_x or a_y equals 1.00, are cooperative interpersonal reward systems.

Results and Discussion

Figure 2 summarizes the essential findings of Experiment 1. Careful study of this figure will clarify the experimental conditions and the findings considerably. Results of subsequent experiments are presented in a similar manner.

The columns of small graphs represent the first two and last two successive blocks of five trials, as labeled across the top. These trial blocks seemed sufficient to demonstrate both (*a*) the direction of response changes during the 50 trials, and (*b*) the stability of the response values at the end of the experiment (comparisons of Trial Block 41–45 with 46–50). Each row of graphs represents one of the five reinforcement systems and is labeled in the first graph of each row. For example, S^r .525–.525 in the second row indicates that reinforcing stimuli were based upon $Z_x = .525\ R_x + .475\ R_y$ and $Z_y = .475\ R_x + .525\ R_y$. The two coordinates, *x* and *y*, for each point plotted are based upon the averages of response magnitudes for five trials performed by S_x and S_y, respectively. A symbol shape (e.g., triangle, circle) is used consistently *within* each row to represent measurements from the same dyad, but similar symbol shapes, among different rows, refer to different dyads. A single symbol always represents responses made by the two members of a *single dyad*, never an average of several dyads.

Within a graph, the two pairs of parallel lines were derived from the equations for Z_x and Z_y. The area between a set of parallel lines is the Good range for either S_x or S_y. The area below and/or to the left of a set of parallel lines is the Poor—Too Little range and the area above and/or to the right is the Poor—Too Much range for an *S*. The combinations of these S^r events for each subarea are shown

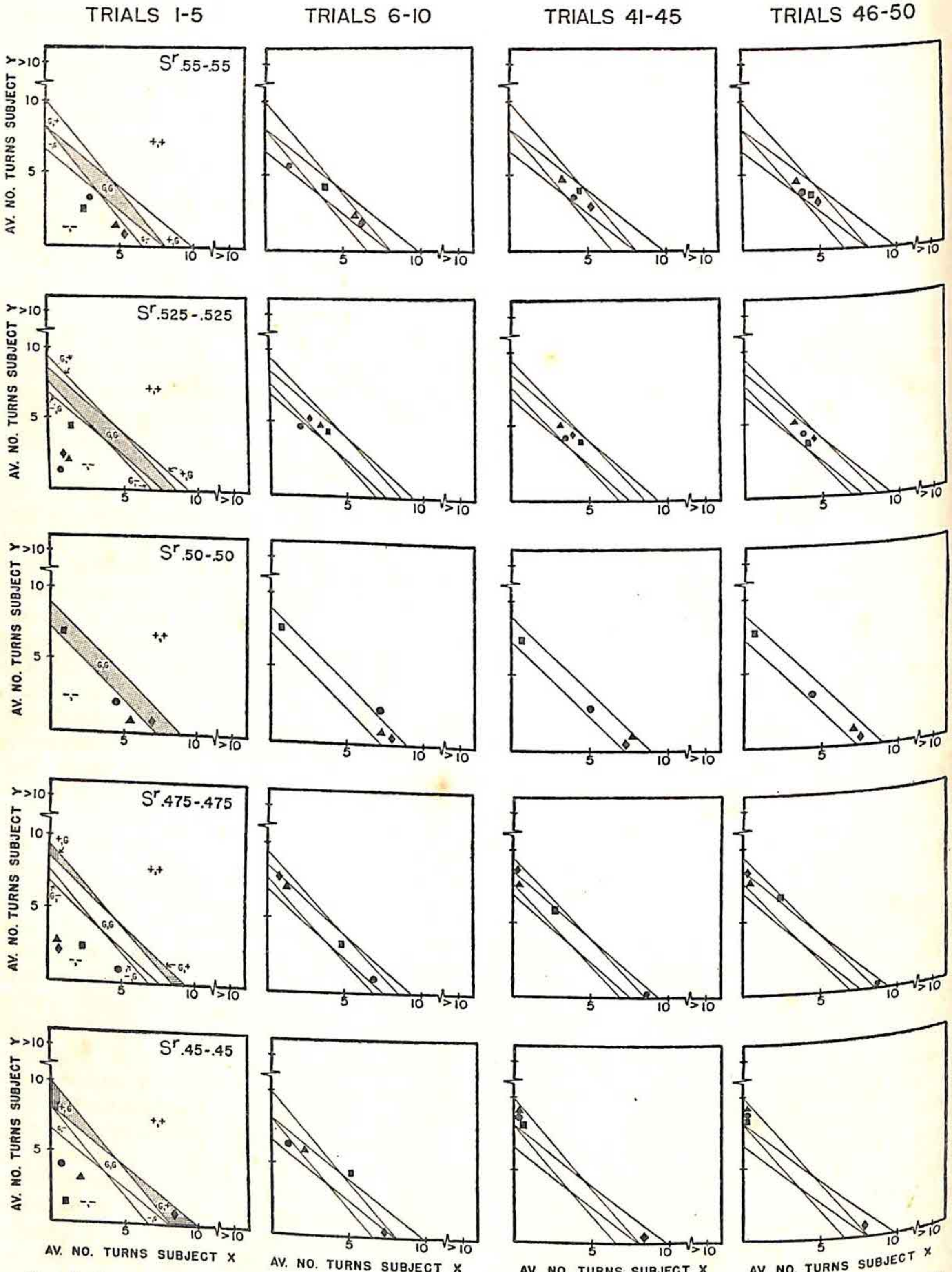


FIG. 2. Summary of the results of Experiment 1. The first two and last two successive blocks of five trials are shown as columns of small graphs and the five reinforcement systems as rows of small graphs. A symbol shape (e.g., triangle, circle) within each row represents the same dyad, but similar shapes among different rows refer to different dyads.

in the first graph of a row. For example, if the two response magnitudes were such as to fall anywhere in the (+, +) area, both *S*s were told Poor—Too Much; or, if the two responses fell into the (G, −) area, S_x was told Good and S_y was told Poor—Too Little; etc.

The shaded areas in the first column of graphs are the areas in which it is predicted that a pair of responses would stabilize over a series of trials and are termed "absorbing areas." The presence of two absorbing areas within a reward system indicates a prediction of response combinations in one *or* the other area, and not that there will be oscillation between these areas.

The absorbing areas are derived from the following simple assumptions:

1. An *S* will tend to decrease his response magnitude following a trial in which he is told Too Much and increase it if he is told Too Little. (This assumption fails to hold, of course, at the extremes, i.e., either when an *S* is giving an extremely small response and is told Too Much, or when he is giving a large response approaching the limit of his speed of turning and is told Too Little.)

2. If *S* is told Good the change in his response magnitude on the following trial is smaller than changes with the other two feedback stimuli (but not zero). (It is important to note that over a series of Good trials, these small response changes would not vary around a constant magnitude. Changes would either occur according to a "random walk" and/or tend to drift in the direction of least effort.)

3. The directions of the response changes by S_x and S_y following a given S^r are assumed to be additive and only the directions of response change are required to predict absorbing area(s). The rate of response changes is ignored in the summations to determine these areas. (Where two areas are predicted or where the predicted area is large, differential initial rates of response change probably determine which of the two areas or which portion of a single area will be the absorbing one. If one is interested in the proportion of trials spent in each area, response rates are also important.)

With these assumptions, a vector of response change may be drawn for each area within a graph. These vectors are all either vertical, horizontal, or at a 45° angle. A few examples are:

−, − ↗
G, − ↑
+, G ←
+, + ↙
G, G −

Having drawn vectors in all the subareas, one can predict the resultant movements of a combination of responses over a series of trials. In the first three rows in Fig. 2, the functions of Z generate a set of vectors converging on the (G, G) area from *all parts* of all other subareas. A response combination occurring in any subarea other than (G, G) would be followed by an S^r combination tending to change the responses toward the (G, G) area only. In the bottom two rows, however, vectors in two subareas, (G, +) and (+, G), do not point exclusively to the (G, G) area. That is, the S^r combination in a small portion of each of these two subareas is expected to change the response magnitude of one *S* but the change should still maintain the response combination in that subarea. Even if response combinations do occur in the (G, G) area during the process, these combinations are not stable over time (Assumption 2, above) unless all parts of all subareas lead to the (G, G) area. The ultimate absorbing areas in the bottom two rows are, therefore, (G, +) and (+, G) subareas and not the (G, G) area. The deeper shading within these two absorbing areas represents a prediction of responses to that part of the absorbing area, i.e., where one *S* is performing extremely small responses and the other *S* is performing responses somewhere within the range designated by the deeper shading. This area of response combinations is termed, for convenience, an "absorbing wall." The absorbing wall exists because there is no provision for "negative responses." In the case of this study, no turns of the knob in the opposite direction are possible. If this direction of turning *were* permitted, the prediction for these two reward systems would not be of the same areas but would be of areas with a similar consequence, i.e., a failure in the optimal development of cooperative action. The important point is not whether or not an absorbing wall exists, i.e., where one *S* can be told Too Much for very small response magnitudes and cannot make negative responses, but rather that there are parametric values of interpersonal reward systems that lead to the breakdown of optimal cooperative action.

Other Z functions, outside the domain of the present study, may generate a set of vectors for which the resultant is not readily apparent and/or for which one would predict continuous movement from one subarea to another rather than the presence of absorbing area(s) or wall(s).[4]

Visual examination of the graphs shows striking accuracy of prediction for almost every dyad in every condition. The only reward system in which there appears to be any deviation from prediction is in S^r .475–.475 (fourth row) and this can probably be attributed to the relatively small absorbing area. Even here, however, there is no radical departure from prediction. The points are all in close proximity to the predicted areas. Comparison of Trial Block 41–45 with Trial Block 46–50 also shows, graphically, the high stability over time of response pairs in all conditions after 40 trials.

Two analyses were performed to establish statistical significance of these observations. The first tested the adequacy of the predictions over all conditions. The second analysis examined the possibility of differentially effective prediction among conditions. Both analyses were limited to data from the last 10 trials. For each of the 20 dyads, a count was made of those trials in which the response combinations fell outside the predicted area. This number was subtracted from the number of trials in which prediction was verified. These 20 difference values (theoretical range: $+10$ to -10) had a mean of $+5.20$ and σ of 5.60. The positive value for the mean indicated more hits than misses on the average. A t for differences against the null hypothesis that there were as many hits as misses yielded a value of 3.883 (*df* 19), significant beyond the .001 level.[5] The second analysis was a simple F test (between and within groups breakdown) in which the same difference scores were compared among the five systems. The null hypothesis was that each of the five conditions was equally predictable as measured by these difference scores. A qualitative comparison of the mean difference for each condition does suggest that S^r .475–.475 is somewhat less predictable than the others, but an F of 1.88 (*df* 4, 15) is not sufficiently large to reject the null hypothesis at the .05 level.

The statistical examinations of the data support the readily observable conclusion that availability of a cooperative situation does not assure optimal cooperative action, i.e., procurement of continuous positive reinforcement for both *S*s on every trial. If we view the five reward systems as points on a continuum, there appears to be a critical value below which *optimal* cooperative action fails to develop. In terms of Experiment 1, both *S*s receive continuous positive reinforcement when the reinforcement value to each is based upon 50% or more of his own response magnitude, i.e., when $a_x = a_y \geq .50$ in the two Z functions. Proportions less than 50% result in partial or no reinforcement to one *S*. Experiment 3 will demonstrate that $a_x = a_y$ is not a necessary condition for either of these effects and will provide a generalization for all reward systems depicted by Fig. 1.

Before additional empirical exploration, it is worth examining the features of these cooperative situations which covary with manipulations of the Z functions. It is obvious from Fig. 2, for example, that the relative sizes of the subareas vary from row to row. And in S^r .50–.50 four subareas are completely absent. However, there is little reason to interpret the results in terms of available areas or their sizes relative to each other. The selection of systems above and below the middle row was such as to control for subarea sizes. Thus, for example, S^r .55–.55 and S^r .45–.45 have comparable subareas and they are of the same size.

It is possible to suggest other interpretations of the results in reward systems that fail to elicit optimal action. It may be that the absorbing areas are valid because of the improbability that *S*s ever encounter the (G, G) outcome. If they did, this interpretation would predict that their responses would stabilize in (G, G) rather than in (G, +) or (+, G). The data fail to support this interpretation. In S^r .45–.45 and S^r .475–.475 there is an average of 7.0 and 19.8 (G, G) outcomes in the first 40 trials. This fact is discernible to some extent in Fig. 2 for Trial Blocks 6–10. In addition, results of a previous study (Rosenberg, 1959) demonstrated very clearly that even if *S*s are first individually trained to a point where they are both receiving continuous reinforcement, i.e., placed in the (G, G) area, the change to a cooperative condition results in gradual deterioration of reinforcement if the relative response weighting is less than S^r .50–.50.

---

[4] Many of the two-person game matrices of current interest to probability learning theorists and to social psychologists appear to result in asymptotic response probabilities in a cell or "area" of less than 1.0, i.e., continuous movement from one cell to another. More quantitatively stated assumptions are required to predict these response probabilities for each area in such interpersonal reward systems (Atkinson & Suppes, 1958; Estes, 1957). It is interesting to note that response *sequences* in these systems rarely receive analytic treatment. The present paper is an attempt in this direction.

[5] This is a rather conservative estimate of the accuracy of the prediction since it fails to credit occurrences in proximal areas, i.e., "near misses" and since predicted areas are very small relative to total area.

Finally, one may explain stability of a nonoptimal absorbing wall in terms of partial reinforcement for one *S* and continuous reinforcement for the second *S*. Since the (G, G) area is adjacent to the (G, +) and (+, G) areas, the (G, G) outcome may occur with sufficient frequency to reinforce S_x or S_y (as the case may be) for very small response magnitudes, and thereby produce relatively stable behavior. This explanation was tested in Experiment 2 by studying more extreme conditions where partial reinforcement was less likely to occur, i.e., where the (G, G) area was not adjacent to (G, +) or (+, G) areas. This extension was combined with an analysis of a second parameter of the reinforcing stimulus, i.e., its informational ambiguity.

EXPERIMENT 2

In essence, the results of Experiment 1 demonstrated that *S*s who are placed in a reward system in which relatively few outcomes may be misleading fail to gravitate toward optimal cooperative behavior in spite of the presence of an optimal outcome and in spite of the occurrence of such an optimal outcome during the learning sequence. This finding suggests that a cooperative situation be constructed using an unordered set of reinforcing events, and hence, no "misleading" outcomes. Such unordered S^r events are less explicit either because (*a*) they are only qualitatively discriminable to *S*, although outcomes do implicitly covary with the quantitative combinations of R's or (*b*) unrewarding outcomes actually contain no information about what kind of response change is needed to bring about the rewarding outcome. The reduction of explicit information by either alternative may retard learning in cooperative situations where explicit information would have resulted in rapid learning. On the other hand, an unordered set of outcomes may produce behavior that yields a larger proportion of rewarding outcomes in cooperative situations where more explicit information about Z values would be misleading. Thus, three sets of reinforcing events varying in informational content about Z were compared in Experiment 2. Information about the direction of nonreinforcement has been studied extensively in individual learning (Stevens, 1951, pp. 1267–1270), but has received scant attention as a parameter in interpersonal learning. Finally, a wider range of *a* values was studied in Experiment 2 in order to assess the limits of generality of results in Experiment 1.

Experimental Conditions

Each dyad was assigned to one of the five reward systems indicated in Fig. 1 by the number 2 in a circle, and to one of the three sets of reinforcing events. Thus, there were 15 experimental conditions (5 × 3) with four dyads in each condition.

In terms of the equations

$$Z_x = a_x R_x + (1 - a_x) R_y$$
$$Z_y = (1 - a_y) R_x + a_y R_y$$

a_x equalled a_y within a dyad. As in Experiment 1, therefore, S_x received the same proportion of R_x as did S_y of R_y in each dyad. The five values of *a* were 1.00, .60, .50, .40, .00. It might be noted that the first combination of values of *a* (1.00, 1.00) is at an extreme which is completely asocial since neither *S* depends upon the other for his reinforcement. It is the only combination in Experiment 2 which fails to fulfill the definition of a cooperative situation.

The three sets of reinforcement events varying in information about Z were:

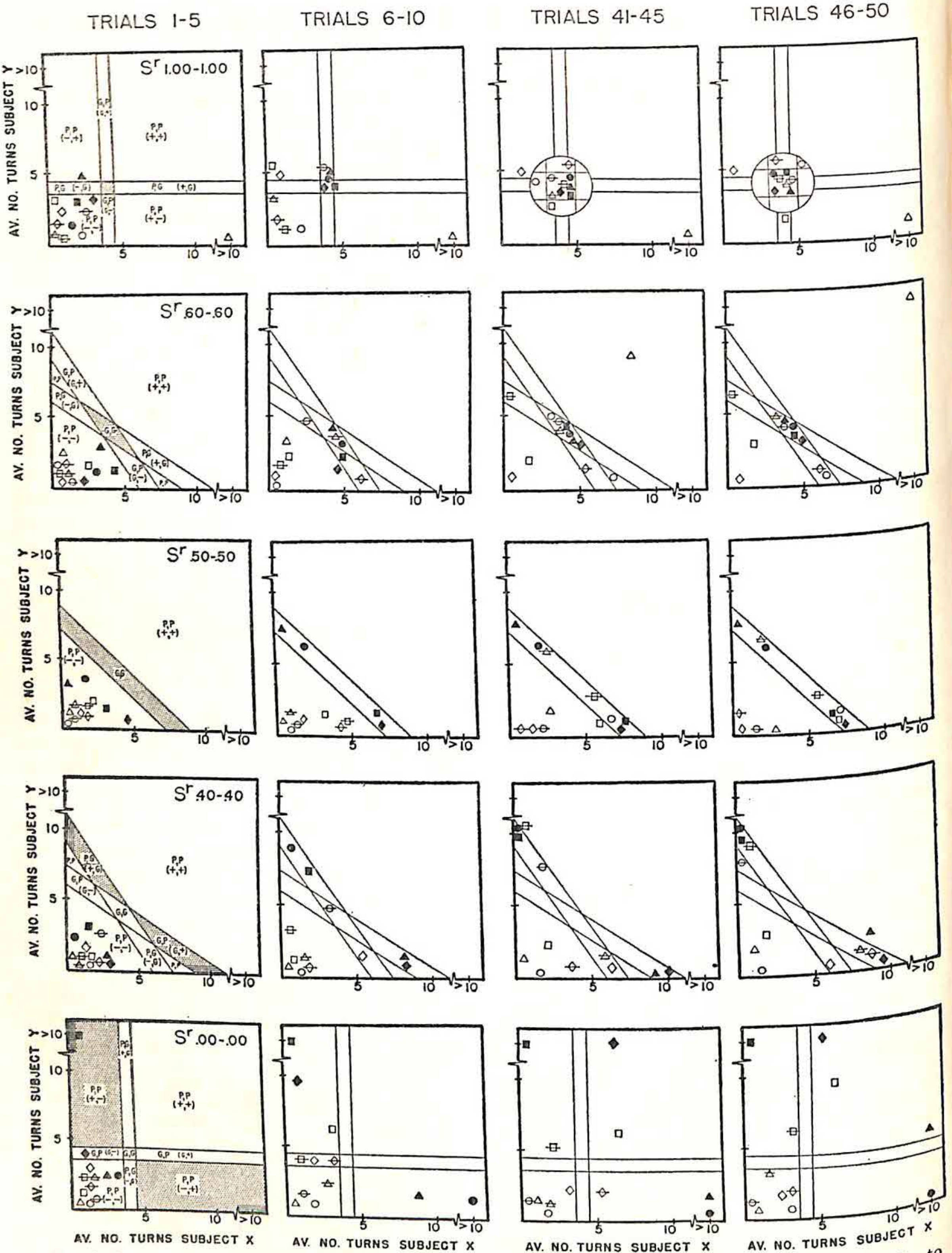


FIG. 3. Summary of the results of Experiment 2 (see explanation of Fig. 2). The filled symbol shapes refer to dyads given explicit information about plus or minus values shown in each subarea of a graph; the open symbols with horizontal bar refer to dyads given ambiguous information about the plus or minus values; the open symbols represent dyads given only "poor" or "good" as outcomes.

1. Poor—Too Much; Good; Poor—Too Little (MGL)
2. Poor; Good; Poor (PGP)
3. Poor; Good (PG)

The first set is the same as that used in Experiment 1. In the second set, Too Much and Too Little were left off, but which of the two Poor outcome lights was given on a trial depended consistently on whether the Z value was above or below the Good range. In the third type, all Z values above and below the Good range were called Poor, and hence only two outcome lights were used, one for Good and one for Poor.

Results and Discussion

Figure 3 is a graphic summary of the results. Format is similar to that employed in Fig. 2. Additional notations were required to distinguish the three sets of reinforcement stimuli. The solid black symbols represent the MGL set (used exclusively in Experiment 1), the symbols with horizontal bar represent the PGP set, and the open symbols represent the PG set of two stimuli. Use of the same symbols within a row indicates successive measurements from the same dyad while similar symbols in different rows refer to different dyads.

Also, each subarea on the graph is labeled with two pairs of coordinates. Coordinates in parenthesis refer to the explicit or implicit values associated with the MGL and PGP conditions, respectively. Coordinates outside of parenthesis refer to a pair of outcomes in the PG condition. The shaded areas were derived from the same assumptions stated in Experiment 1, although the validity of these assumptions for the PGP set and certainly for the PG set is questionable.

An analysis was first performed to ascertain the differential validity of these assumptions on the last 10 trials for all 15 conditions. A factorial design was employed containing two criteria of classification: (*a*) five reward systems and (*b*) the three sets of reinforcement events varying in informational content. The dependent variable for each dyad was simply the number of trials between 41 and 50 in which the response combination occurred in the preducted area minus the number of occurrences in all other areas. The analysis is summarized in Table 1 and indicates a significant *F* only among the three reinforcement sets varying in information. The graphs in Fig. 3 illustrate very clearly the nature of these differences. The assumptions for deriving the shaded area are least valid for the PG condition. There are almost no dyads in this condition whose behavior conform to prediction. The assmptions are more valid for the PGP conditions but still fall short of the very consistent findings for MGL condition. Deviation from prediction does occur in the MGL condition also, particularly near the extreme where the *a* values are small, i.e., where cooperative action is not predicted to develop optimally. In S^r .00–.00, where it was predicted that neither *S* would receive reinforcement, the four dyads conform to prediction in the early trials, e.g., 6–10, but by the end of the experimental session two dyads have moved out of the predicted areas. Also in S^r .40–.40, one dyad changed responses so as to move out of the predicted area during the last five trials. These response changes out of predicted areas, interestingly enough, do not result in more effective cooperative action, at least during the number of trials studied. These findings for small *a* values lend support to the hypothesis suggested in discussion of Experiment 1, i.e., that a nonoptimal absorbing area may depend for its long-term stability on partial reinforcement of the small response magnitudes by one *S*. A significant interaction of reward system with information content, i.e., 1 × 2 in Table 1, could support this interpretation more fully. The absence of interaction fails to refute, however, the generality of the assumptions used to derive predicted subareas. More *S*s and more trials may be required to refute the validity of the assumptions for extreme

TABLE 1

ANALYSIS OF VARIANCE OF THE LAST 10 TRIALS OF EXPERIMENT 2 USING THE DIFFERENCE SCORES (FREQUENCY OF PREDICTED AREAS MINUS FREQUENCY OF OTHER AREAS) AS THE DEPENDENT VARIABLE

Source	*df*	Mean Square	*F*
Reward System	4	20.83	.70
Reinforcement Sets Varying in information	2	856.06	28.96*
1 × 2	8	27.73	.94
Within Cells	45	29.62	(error term)

* $P < .001$.

reward systems involving the set of outcomes, MGL.

Accurate prediction for the PGP and PG conditions obviously cannot be completely derived from these assumptions. The PGP condition results in behavior that conforms, in part, to the predictions. Apparently the early trials provide a sufficiently consistent relationship between changes in response magnitude and presentation of the two Poor outcomes so that *S*s learn to "impose" an appropriate ordering of these two Poor outcomes, i.e., discriminate between them. The particular reward system may be such as to continue to reinforce this discrimination of the two Poor outcomes. However, in other reward systems (bottom two rows) the decrement in reward frequency may result in the extinction of this implicit ordering behavior. Whether, at this point, the *S* reverses the values, i.e., reverses his previous discrimination, or simply lumps the two Poor outcomes, i.e., the discrimination is extinguished, or is inconsistent is difficult to infer from these data. In any case, however, the frequency of reward does not increase appreciably over time in the conditions represented by the bottom two rows. There is, simply, a lower frequency of the very large or very small response magnitudes that are found in the MGL condition.

The validity of the predicted areas for the PG condition is nil. This finding is not surprising since the assumptions from which these areas are derived require informational components in the reinforcement not present in the PG condition. The *S* is given no information about response magnitude, and hence there is no reason to assume either a consistent increase or a consistent decrease in response magnitude following a Poor outcome, i.e., the first assumption for an MGL condition would be discarded. One could retain the two remaining assumptions, however, which are essentially that an *S* will tend to make a larger change in response magnitude (sometimes increasing it and sometimes decreasing it) following a Poor outcome than following a Good outcome and that the directions of response changes by the two *S*s are additive. With these assumptions, it is possible to derive a number of interesting statements that contrast with those made for the other two types of conditions. First, one would predict that in the *cooperative* reward systems in this study, the most stable subarea, if any, would be the (G, G) area. That is, if a response combination falls into the (P, G) areas, S_x will make large response changes. Similarly, if the combination is (G, P), S_y will make large response changes. A (P, P) outcome will yield large response changes by both *S*s. The sizes of these areas relative to the (G, G) area should, therefore, be *one* determinant in the frequency of reward for *both* *S*s. Hence, S^r .50–.50 might be expected to produce more Good outcomes than adjacent conditions since it has the largest (G, G) area and no (P, G) or (G, P) areas. It is the one condition in which both *S*s make large response changes or neither does.[6] This seems to be the effect for at least two of the four dyads in this condition. By contrast, neither the adjacent conditions S^r .60–.60 nor S^r .40–.40 reward systems produce nearly such a high proportion of reward for both *S*s.

The relevance of the Z functions themselves in the PG condition is also derived from this reasoning. We have assumed that (P, G) and (G, P) are unstable to some extent because either S_x or S_y, respectively, makes large response changes. The extent to which these outcomes *are* unstable depends in turn on the pair of Z functions. Thus, if the outcome is (P, G) and the response of S_x received a large weighting in the reinforcement of S_y, it would be expected that S_y would soon fail to receive positive reward. This in turn increases the size of response change of S_y. On the other hand, Z functions that make one or both *S*s relatively independent of the other would produce a greater number of Good outcomes for one or both *S*. This is discernible in the extreme condition S^r 1.00–1.00 where both *S*s are independent of each other. Nevertheless, S^r .50–.50, which appears to clearly divide optimal from nonoptimal reward systems in the MGL condition, is not the critical point in the PG condition.

---

[6] It is interesting to note that the S^r .50–.50 is the only reward system which satisfies some definitions of a cooperative situation, viz., when one receives reward, all do, when one fails to receive reward, no one does (Lindzey, 1954, p. 810). One may speculate about the possibility that this restrictive definition evolved because cooperative action develops best here with equivocal outcomes and that the tasks typically used in the past for cooperative studies do contain more equivocal reinforcing events than those in Experiment 1.

It would be interesting to pretrain both *S*s individually so that their response combinations fell into the (G, G) area and then change them to a cooperative situation. Because of the response drift, and on the basis of this experiment, it is quite probable that a (G, P), (P, G), or (P, P) condition would soon occur and that return to (G, G) would then be as difficult for such a dyad as for one which had received no preliminary individual training.

In summary, then, using a PG set of outcomes, one would probably predict that no subarea is as stable as those found in more informative outcomes. In general, one would predict that the (G, G) area is most stable, the (P, G) or (G, P) areas next most stable, and (P, P) areas least stable. The frequency of occurrence of responses in any of these areas depend upon (*a*) the relative sizes of these areas and (*b*) the degree of independence between persons as determined by the pair of Z functions.

Quantitative assumptions are required to make more exact predictions about outcome stability as a function of these experimental parameters. Qualitatively, it seems surprising, nevertheless, that the wide variations in the Z functions in this study (short of complete autonomy) result in such an ineffective procurement of positive reinforcement for both *S*s over all 50 trials in the PG condition. There is some reason to think that this poor performance is, in part, a function of the type of *S* used in this study. It was observed that in many dyads both members failed to demonstrate much exploratory behavior over the entire 50 trials. They failed to increase their response magnitudes sufficiently to get out of

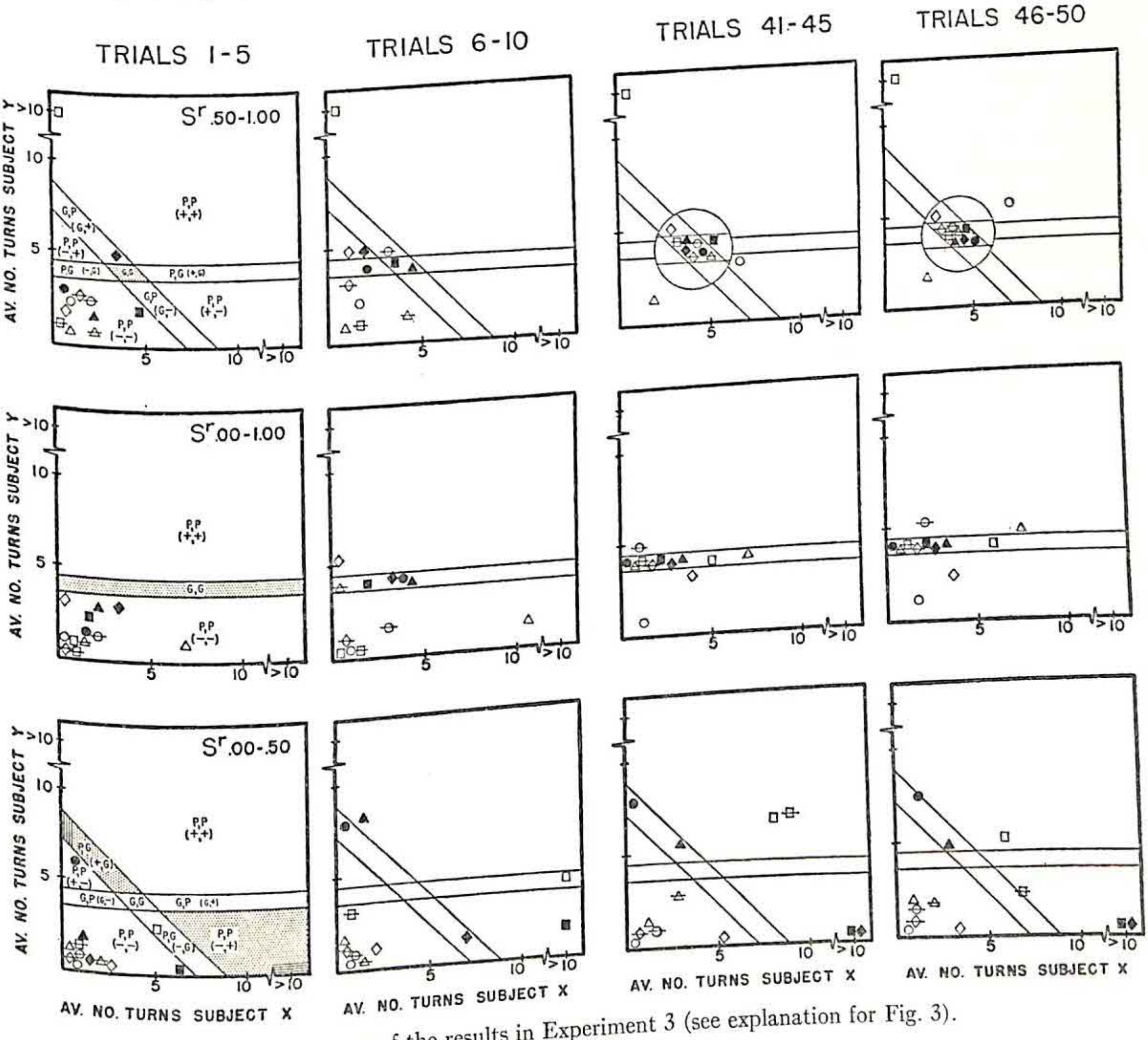


Fig. 4. Summary of the results in Experiment 3 (see explanation for Fig. 3).

the lower left (P, P) area. This depressed activity may be inherent in the nature of the sample of *S*s. They were all persons who had just entered Air Force service and hence were in a very strange and stressful environment generally. It is unknown whether other, more relaxed *S*s may have simply elevated the total frequency of reward over all reward systems, or whether the effects of different reward systems would have been more clearly distinguishable using the PG outcomes with the more "relaxed" *S*s.

EXPERIMENT 3

The purpose of the final experiment is best understood by reference to Fig. 1, the graphic representation of the combinations of Z functions for each dyad. The results of the first two experiments indicate that a predicted subarea for the most explicit form of reinforcement stimulus, MGL, is unrewarding to one or both *S*s when a reward system is imposed in which both receive as feedback less than half of their own response variance in the S^r and when $a_x = a_y$. This is the lower half of the hypotenuse of the triangle in Fig. 1. By extrapolation, it may be demonstrated that these two restrictions are not necessary conditions for predicting an absorbing area containing an unrewarding outcome for one *S*. In fact, a nonoptimal absorbing area will be predicted even though one *S* may receive a *majority* of his response in his reinforcement as long as

$$a_x + a_y < 1.00$$

This equation defines the shaded area in Fig. 1. When $a_x \neq a_y$ the effect is to increase the probability of unrewarding outcomes to the *S* assigned to the smaller *a* value and decrease the probability of unrewarding outcomes to the *S* assigned the larger *a* value.

The purpose of this experiment is to demonstrate these predictions empirically by selecting three reward systems not on the hypotenuse of the triangle where

$$a_x + a_y > 1.00$$
$$a_x + a_y = 1.00$$
$$a_x + a_y < 1.00$$

Experimental Conditions

Each dyad was assigned to one of the three degrees of S^r information, i.e., MGL, PGP, PG, and to one of the three reward systems shown in Fig. 1 by the number 3 in a circle. This made a 3 × 3 design with four dyads in each cell.

The sum of values for a_x and a_y were 1.50, 1.00, and .50. The *a* values were, respectively, (*a*) .50, 1.00; (*b*) .00, 1.00; (*c*) .00, 50.

TABLE 2

ANALYSIS OF VARIANCE OF THE LAST 10 TRIALS OF EXPERIMENT 3 USING THE DIFFERENCE SCORES (FREQUENCY OF PREDICTED AREAS MINUS FREQUENCY OF OTHER AREAS) AS THE DEPENDENT VARIABLE

Source	*df*	Mean Square	*F*
Reward System	2	85.77	4.92*
Reinforcement Sets Varying in information	2	341.77	15.60***
1 × 2	4	95.45	5.47**
Within Cells	27	17.44	(error term)

* $P < .05$.
** $P < .01$.
*** $P < .001$.

Results and Discussion

Figure 4 is a graphic summary of the results. The same notation was used here as in previous figures.

Qualitative inspection supports the validity of predictions for the MGL condition. The two other types of S^r information show less tendency to produce response pairs in the predicted areas, and again, the PG condition is least predictive. Note also that the PG condition is in general somewhat less sensitive to the different reward systems than are the other two informational conditions.

An analysis of variance of the results for the last 10 trials is summarized in Table 2 and corroborates the significant differences among the three reinforcement types. Significant effects attributable to different reward systems and to interaction were also obtained. Figure 4 suggests the nature of these effects. The PGP and, to some extent, the PG conditions show better predictability when (G, G) is the predicted outcome than when (P, G) or (G, P) is predicted. This result is in line with the discussion of these conditions in Experiment 2.

GENERAL DISCUSSION

There appears to be a growing interest in the simplification and increased control of laboratory situations employed to study interpersonal phenomena. This trend stems, in part, from the analyses of interpersonal behavior based upon an S-R system. (Lindzey, 1954, pp. 77–87; Logan, 1955; Miller & Dollard, 1941; Sears, 1951.) These analyses usually include concepts and principles of learning, and the empirical studies almost invariably employ laboratory situations that are highly controlled with regard to the stimulus and response elements of an interpersonal event (e.g., Miller & Dollard, 1941; Rosenberg & Hall, 1958; Rosenberg, 1959; Sidowski, Wyckoff, & Tabory, 1956). Some writers, whether they apply an S-R system or not, employ these highly controlled situations for the application and development of rather formal theoretical systems, e.g., mathematical models of interpersonal events (Atkinson & Suppes, 1958; Estes, 1957; Hays & Bush, 1954).

The present study is concerned with the problem of organizing a set of interpersonal variables on a descriptive S-R level and in refining experimental techniques rather than in quantitative theory construction. The descriptive approach, however, is not aimed at the proliferation of experimental variables without regard to some parsimonious grouping of experimental operations. The present study, for example, would appear to provide generalization for a number of Z functions in addition to the simple additive ones used here. It would appear to matter very little, for example, whether the functions for S_x and/or S_y were curvilinear rather than linear. If one could determine unequivocably that absorbing area(s) existed for a pair of Z functions (and this is possible for a variety of nonlinear functions), it would matter little that these functions were nonlinear.

There are combinations of functions, however, for which no absorbing area(s) can be predicted from the assumptions in this paper. For example, if one used the function $Z = aR_y - bR_y$ for one *S* and $Z = a'R_x + b'R_y$ for the second *S*, one would be able to predict from the assumptions used in this study only that R_x and R_y would change continually from trial to trial so as to "orbit" around the (G, G) area. There is no way of predicting from present assumptions whether the orbit would become smaller and hence converge on the (G, G) area, become larger, change shape, etc. Experimental results using this interesting pair of general functions would presumably permit some generalization to a variety of similar combinations of functions. It is interesting to note, parenthetically, that the $Z = aR_x - bR_y$ function for one *S* in combination with a function describing individual discrimination learning by a second *S* resembles the imitation paradigm of Miller and Dollard (1941) and that an absorbing area does exist for this reward system in the (G, G) area.

Another interesting set of function pairs is that in which no (G, G) area exists. That is, the Z functions do not intersect anywhere on the graph. This specification would define a competitive task, if only one trial were permitted. Over a series of trials, however, such a combination of functions would fall within the definition of a cooperation. It would not be surprising, however, to find dyads unable to develop optimal cooperative action nor even to procure an optimal sum of individual reward frequencies although the situation permitted the occurrence of a rewarding outcome to one or the other *S* on each trial. Studies that combine interests in learning and in game matrices usually employ examples of this set of function pairs.

A second class of variables subject to empirical analysis and generally independent of the Z functions, are those termed "awareness". While these variables have received considerable prominence in social psychology (some social psychologists would reject a situation as interpersonal if *S*s were not "aware" of social linkages), there has been little systematic attempt to formulate a parametric approach for this class of operations. When "awareness" is manipulated by *E* directly, it usually takes on only two values—an *S* is or is not told something. (Neimark & Rosenberg, 1959; Sidowski, 1957) Some experimental tasks, e.g., face-to-face, are so constructed that extensive manipulation of awareness is not possible. Other tasks such as the one in this study permit the *E* to assign many "values" to awareness. These values may range from telling each *S* nothing about the interpersonal linkage during the initial instructions, to a complete description for both *S*s of the functions by which their responses are to be combined. Note also that the task in this study permits the *E* the choice of delivering to each *S* rewards of which the other *S* may or may not be made aware from trial to trial, and thereby extended the concept beyond the instructional differences.

The implications of the present study would appear to extend beyond that of a knob-turning task. Pilot replications of aspects of this study by the author using other trial-and-error tasks, e.g., time estimation, size estimation, suggest ready generalization of results to these tasks. One would also expect, for example, some generalization to the autokinetic response. While it is premature to outline implications of the present results to life situations, it seems likely that the interpersonal reward systems imposed by laboratory situations can be coordinated with systems of reward imposed upon a dyad by some natural or mechanical environments—and perhaps can be eventually coordinated to reward systems imposed by the behavior repertoires of the interactors.

SUMMARY

Reinforcement parameters for a dyadic cooperative situation were defined and varied in a series of three experiments. A cooperative situation was defined as one where (*a*) there exist at least two outcomes for each *S*, (*b*) one of which is a rewarding state of affairs, (*c*) reward delivered to one person does not preclude its availability for the other, and (*d*) the outcomes to *both* persons are, in part or entirely, a function of the behavior of the other person. The task was the same in all three experiments. Each *S* was required to perform the simple motor response of turning a small concealed knob in a series of 50 trials. A reinforcing stimulus was delivered to each *S* separately after each trial. All conditions either satisfied the definition of a cooperative situation or were a social (wherein reinforcement was contingent upon individual behavior alone). The degree to which optimal cooperative behavior, i.e., procurement of maximum reinforcement by both *S*s developed in these situations, was studied as a dependent variable.

Experiment 1 was composed of five cooperative reward systems. These systems varied in the extent to which the response magnitude of one *S* of a dyad determined the reinforcement stimulus for the other *S* and vice versa. That is, the reinforcement stimulus, Good, was delivered to an *S* on trials when the Z value in the following equation fell within a predetermined range.

$$Z = aR_x + (1 - a)R_y$$

where R_x and R_y are response magnitudes of the two *S*s and where a is assigned a value between zero and one for each *S*. Predictions were made concerning the degree to which cooperative action would develop as a function of these relative values of a for each *S*. It was predicted that in three reward systems, where *both* *S*s' reinforcement stimulus was contingent upon a fraction equal to or greater than .50 of his own response magnitude, optimal cooperative behavior would develop quickly and would be maintained consistently. In the two remaining reward systems, however, where both *S*s' reinforcement stimulus was contingent upon less than .50 of his own response magnitude, optimal cooperative behavior was not predicted. The results clearly substantiated these predictions.

Experiment 2 extended the range of relative response weightings to its full range, i.e., a values from zero to one were used for different sets of dyads. This extended range tested the generality of the predictive assumptions used in Experiment 1. Predictions were substantiated for this extended range of weightings. The weighting, .50, was again demonstrated to be the critical point, i.e., when reinforcements to both *S*s were weighted by less than .50 of their own responses, optimal cooperative action failed to develop.

Informational content concerning the size of Z was varied in Experiment 2. It was found that predictions about *S*s' behavior were less valid as the information to *S* about Z became more equivocal and that the total frequency of positive reinforcement to either *S* dropped considerably under these conditions of reduced information.

Experiment 3 investigated further extensions of the response weighting parameter so as to demonstrate the generality of the predictive assumptions for all *combinations* of weightings. Experiments 1 and 2 were limited to combinations of weightings where the response of one *S* was weighted the same in his reinforcement stimulus as that of the second *S* in *his* reinforcements. Experiment 3 demonstrated when

one *S*'s response weighting in his reinforcement is different from the second *S*'s response weighting, optimal cooperative action fails to develop where the sum of these two weights is less than 1.00, and hence such failures are not limited to reward systems where *both* *S*s' responses are weighted less than .50.

These experiments support the conclusion that the combination of response weightings in the reinforcements to each *S* is an important parameter in the development of cooperative action. The amount of information concomitant to the presentation of reinforcement to each *S* also appears to play an important role in the development of cooperative action and in the frequency of reinforcement procured by both *S*s over a series of trials.

The status of the present study as a contribution to a descriptive system of interpersonal behavior was briefly discussed. Other experimental parameters were suggested and the implications of these parameters for analysis of interpersonal situations were sketched.

REFERENCES

Atkinson, R. C., & Suppes, P. An analysis of two-person game situations in terms of statistical learning theory. *J. exp. Psychol.*, 1958, **55**, 369-378.

Deutsch, M. A theory of cooperation and competition. *Hum. Relat.*, 1949, **2**, 129-151.

Estes, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609-617.

Gragg, D. B., & Douglass, H. J. Conversion tables for selected Airman Classification Battery scores and comparable scores on other selected service and civilian tests. *USAF Pers. Train. Res. Cen. Develpm. Rep.*, 1956, No. AFPTRC-TN-56-89.

Hall, R. L. Group performance under feedback that confounds responses of group members. *Sociometry*, 1957, **20**, 297-305.

Hays, D. G., & Bush, R. A study of group action. *Amer. Sociol. Rev.*, 1954, **19**, 693-701.

Lindzey, G. (Ed.) *Handbook of social psychology.* Vols. 1 & 2. Cambridge: Addison-Wesley, 1954.

Logan, F. A. *Behavior theory and social science.* New Haven: Yale Univer. Press, 1955.

Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

Mintz, A. Non-adaptive group behavior. *J. abnorm. soc. Psychol.*, 1951, **46**, 150-159.

Neimark, Edith D., & Rosenberg, S. The effect of "social" discriminative cues on probability learning. *J. exp. Psychol.*, 1959, **58**, 302-311.

Rosenberg, S. The maintenance of a learned response in controlled interpersonal conditions. *Sociometry*, 1959, **22**, 124-138.

Rosenberg, S., & Hall, R. L. The effects of different social feedback conditions upon performance in dyadic teams. *J. abnorm. soc. Psychol.*, 1958, **57**, 271-277.

Sears, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist*, 1951, **6**, 476-483.

Sidowski, J. B., Wyckoff, L. B., & Tabory, L. The influence of reinforcement and punishment in a minimal social situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 115-119.

Sidowski, J. B. Reward and punishment in a minimal social situation. *J. exp. Psychol.*, 1957, **54**, 318-326.

Stevens, S. S. (Ed.) *Handbook of experimental psychology.* New York: Wiley, 1951.

Tinbergen, N. *Social behavior in animals.* New York: Wiley, 1953.

(Received January 29, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

ACHIEVEMENT RELATED CHARACTERISTICS OF MANIC-DEPRESSIVES[1]

JOSEPH BECKER[2]
Duke University

THE general purpose of the present study was to attempt a more objective and rigorous testing of hypotheses about manic-depressive character derived primarily from the clinical research of Cohen, Baker, Cohen, Fromm-Reichmann, and Weigert (1954).[3] Their descriptions of manic-depressive character were reformulated in terms of personality types which have emerged from experimental investigations of human motivation. The Achievement concepts of McClelland and his associates (McClelland, Atkinson, Clark, & Lowell, 1953; De Charms, Morrison, Reitman, & McClelland, 1955) provided a particularly appropriate framework for examining certain aspects of manic-depressive character. The formulations of Cohen et al. attached considerable significance to behaviors and attitudes related to achievement and conformity, which are also of central importance to McClelland's concepts.

According to these writers (Cohen et al., 1954, pp. 113–119), early parental indulgence was abruptly supplanted by approval conditioned upon the fulfillment of parental demands for rigidly conforming behavior and high achievement. This change in the nature of parental demands resulted in rigid, impersonal standards being foisted upon the premanic-depressive child by his parents in order to use him as an instrument for raising the family's social prestige. These standards of achievement and conformity were determined by the parents' (usually the mother's) conception of what the neighbors would think proper and appeared to be severely enforced.

In adulthood (Cohen et al., 1954, pp. 119–121), the manic-depressive was caught in the dilemma of strongly inculcated achievement values based on the need to placate others, and and fear of alienating others through his success. Achievement was valued primarily as a means toward the end of gaining others' approval. Excellence of performance or independent activity as satisfying in itself was subordinated to the belief that success was attainable only through manipulating others, not through one's own efforts. So excessive is the dependence of the manic-depressive on the approval of others that he will abandon even his rigidly conventional opinions in favor of the perceived values of the needed other, particularly if the other is seen as an authority figure.

McClelland and his associates identified two relatively independent types of achievement orientation. Need Achievement is said to characterize persons who are concerned with living up to an internalized standard of excellence. Their satisfaction is in doing a job well rather than in the enjoyment of the end product. As children, high need Achievers were subjected to more early independence demands and had fewer restrictions imposed upon them than did low need Achievers (Winterbottom, 1953). In contrast, value Achievement characterizes individuals who have responded to excessive parental verbal emphasis on achievement striving, by a conscious valuing of achievement. DeCharms et al. (1955) in speculating about the origins of value Achievement state that persons "who describe themselves as ambitious and achievant may do so for defensive reasons; they have perhaps been under some authoritarian pressure from their parents to be ambitious and the resultant motive which has originated in external sources shows itself primarily as a fear of being unsuccessful" (p. 419). Whereas high need

---

[1] This paper is based upon a doctoral dissertation submitted to Duke University. The writer wishes to express his appreciation to Charles D. Spielberger, the dissertation chairman, and Joseph B. Parker, Chief of Psychiatry VA Hospital, Durham, North Carolina, for their generous assistance.

[2] Now at Umstead State Hospital, Butner, North Carolina.

[3] This investigation was based on a research project which studied 12 manic-depressives who were being treated by intensive psychoanalytic therapy. The research group was comprised of staff members of the Washington School of Psychiatry who met in bimonthly, three-hour seminar sessions for approximately four years.

Achievement has been shown to be positively related to better performance on tasks involving "internalized standards of excellence" and negatively related to conformity and authoritarianism (Brown, 1953; McClelland et al., 1953; Wendt, 1955), high value Achievement is unrelated to performance involving internalized standards of excellence and is positively related to conformity behavior and authoritarian attitudes (DeCharms et al., 1955).

According to the conceptualization of Cohen et al. (1954), manic-depressives and high value Achievers seem to share a common defensive reaction pattern consisting of outwardly adopting the prevailing values of others in order to placate and win needed approval from them. Dependency on others for guidance and approval is manifested in the behavior of both manic-depressives and high value Achievers through conventional attitudes, stereotyped social interactions, and conformity. However, Cohen et al. have also asserted that premanic-depressive children frequently assumed early responsibility, and as adults often became financially successful. Such behavior would not be inconsistent with what might be expected of McClelland's high need Achievers. Furthermore, Arieti (1957) has implied that manic-depressives are like Riesman's inner-directed character type (Riesman, Glazer, & Denny, 1950), a type with which McClelland (1955) has also identified the high need Achiever.

Since phenotypical characteristics and the socialization process of manic-depressives seemed to resemble those of high value Achievers more closely than those of high need Achievers, the hypotheses of this study predicted that manic-depressives would perform like high value Achievers, as compared to a nonpsychiatric control group, on a number of tasks which differentiate need and value Achievement performance.

The following predictions were made:

Hypothesis 1: Manic-depressives describe themselves as more ambitious and achievant than the controls.

Hypothesis 2: Manic-depressives subscribe more strongly to authoritarian ideology than the controls. Both high authoritarians and manic-depressives have been reported to share such characteristics as intolerance of ambiguity, social imperceptiveness, stereotypy and rigidity in behavior and attitudes, and strong conformity tendancies (Titus & Hollander, 1957; Cohen et al., 1954).

Hypothesis 3: Manic-depressives profess a more traditional ideology with regard to family roles and statuses than the controls.

Hypothesis 4: Manic-depressives are not significantly different from the controls in strength of need Achievement phantasy.

Hypothesis 5: Manic-depressives are not significantly different from the controls in their achievement on performance tasks under either task or achievement oriented conditions.

If, however, the manic-depressives more closely resemble high need Achievers, then on the operational measures: (*a*) the manic-depressives would not be expected to differ from the controls on the value Achievement scale; (*b*) the manic-depressives would be expected to score lower on the California Fascism scale, and the Traditional Family Ideology Scale; and (*c*) the manic-depressives would be expected to score higher in need Achievement phantasy and the performance tasks.

METHOD

Subjects

The subjects (*S*s) were 24 remitted manic-depressives and 30 nonpsychiatric controls. All *S*s were male white veterans, aged 25 to 65, with at least sixth-grade reading ability; most were born and reared in the Southeastern United States. Remitted manic-depressives were used in this investigation in order to study manic-depressive character undistorted by the confounding effects of severe affective disturbance and treatment regimens. These cases were carefully selected on the basis of Kallman's (1953) definition of manic-depressives as "cyclic cases which show periodicity of acute, self-limited mood swings before the fifth decade of life and no progressive or residual personality disintegration before or after such episodes" (p. 130), the absence of Lewis and Piotrowski signs (Lewis & Piotrowski, 1954), and judgments of clinical remission discussed in detail elsewhere (Becker, 1958). All of the manic-depressive patients had been at the Durham VA Hospital some time during the past five years.

TABLE 1

COMPARISON OF MANIC-DEPRESSIVES AND CONTROLS ON MATCHING VARIABLES

Measure	Manic-Depressives (N 24)			Controls (N 30)			U	p[a]
	Median	Mean	SD	Median	Mean	SD		
Age	41.50	44.13	10.56	38.50	43.17	12.10	344	.78
Education	12.60	12.13	2.91	11.50	12.17	3.66	343.5	.77
Ohio Literacy	36.50	36.50	5.00	37.67	36.50	5.98	382.5	.70

[a] A probability value of $p > .05$ (two-tailed tests) was set as the level of significance required for rejection of the null hypothesis in all of the statistical procedures of this study.

The control group consisted of 30 *S*s who were free of psychosomatic symptoms, chronic physical disorder, and psychiatric illness. The control *S*s were comparable to the manic-depressives for age, education, and reading ability. The medians, means, and standard deviations of the experimental and control groups on the equated variables are presented in Table 1.

Since the data for age, education, and literacy were either markedly skewed or rectangularly distributed, the Mann-Whitney *U* Test was used to determine whether the groups had been drawn from the same population with regard to the control variables. There were no significant differences between manic-depressives and controls on any of these variables.

Procedure

Nine experimental tasks were administered to each *S* in an individual testing session which required approximately 90 min. The following orienting instructions were given in a conversational manner:

> The purpose of the tasks you're about to take is to allow evaluation of some important attitudes and behaviors that seem to have a lot to do with how a person gets along outside the hospital. We want to compare people who have had some emotional difficulties, but recovered or improved, with others who have never been hospitalized for this reason.

The intent of these instructions was to stimulate interest in the tasks without unduly arousing an achievement or affiliative orientation.

The experimental procedures were introduced by asking the *S*s to read the instructions for the Hildreth Mood Scale (Hildreth, 1946) and to fill it out. They were then administered a four-card need Achievement measure ("Inventors," "Boy in Checkered Shirt," "Trapeze Artists," and "Architects") which deviated from the standard procedure in that *S*s were tested individually rather than in a group and the usual 5-min. time limitation per story was not imposed. Most *S*s, however, finished the individual stories within a 5-min. period.

After completing the need Achievement measure, the *S*s were given the Ohio Literacy Test (Foster & Goddard, 1924).

Next, the three attitude scales—value Achievement scale (DeCharms et al., 1955), F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), and TFI Scale—(Levinson & Huffman, 1955), were administered together. Since instructions for each of these scales are similar, only one set of instructions was given. After the attitude scales were completed, the IPIT was given.[4]

The two performance measures were the final tasks administered. The Xs-in-Circles task (Atkinson & Reitman, 1956) was given with task oriented instructions, followed by the Additions task (Williams, 1955). Achievement oriented instructions were used with the latter task, stressing speed, accuracy, and competition.

The finding that the manic-depressives did not differ significantly in affective state and that as a group they were not more affectively variable at the time of testing than the controls supports the clinical judgment that the manic-depressive *S*s were in remission. This finding also supports the contention that any obtained differences on the experimental measures are not primarily due to differences in momentary affective state.

The medians, means, and standard deviations of the experimental measures are reported in Table 3. The differences between the manic-depressives and the controls in the experimental measures were evaluated by *t* tests when the data were approximately normally distributed and by nonparametric Mann-Whitney *U* tests when the data departed markedly from a normal distribution. The findings relative to the hypothesized performance of manic-depressives as value Achievers were as follows:

Hypothesis 1: The scores for the manic-depressives on the value Achievement scale were significantly ($U = 214$, $p < .02$) higher than those of the controls. This finding confirms Hypothesis 1; the manic-depressives placed a strong, positive, conscious valuation on achievement.

Hypothesis 2: Manic-depressives obtained a significantly ($t = 2.05$, $p < .05$) higher mean score than controls on the F scale. This finding confirms Hypothesis 2; the manic-depressives performed in a manner similar to that of high value Achievers on the F scale. High scorers on this scale tend to be characterized by rigidly conventional authoritarian attitudes, conformity, intolerance of ambiguity, and social imperceptiveness.

Hypothesis 3: The mean score of the manic-depressives on the TFI scale was significantly ($t = 2.74$, $p < .01$) higher than was that of the controls. This confirms Hypothesis 3; the manic-depressives performed in a manner similar to that which would be expected of high value Achievers. High scorers on the TFI scale are said to be characterized by attitudes of submission to authority, extreme emphasis on discipline, and rigidly defined roles and statuses for family members.

Hypothesis 4: The distribution of scores for the manic-depressives on the need Achievement measure was not different from the distribution of the controls ($U = 368$). The fact that manic-depressives do not score as would be expected if they were high need Achievers is not inconsistent with their being high value Achievers and supports Hypothesis 4.

Hypothesis 5: The scores of the manic-depressives were not different from those of the controls on either the Xs-in-Circles Task ($U = 372$) or the Additions Task ($U = 309$). The manic-depressives did not perform more rapidly under either the "task" or "achievement" oriented instructions. If the manic-depressives were high need Achievers, they would be expected to score higher on either one or both of these tasks. The finding that manic-depressives did not score higher is not inconsistent with their performing like high value achievers and supports Hypothesis 5.

---

[4] Because of the difficulty in generating hypotheses within a need or value Achievement framework, data for this measure are not reported in this study.

RESULTS

The Hildreth scale served the twofold purpose of checking the reliability of the psychiatric judgment of clinical remission and affording a basis for evaluating the effect of mood state on the experimental measures. The means and variances of the mood scale scores for the manic-depressives and controls are presented in Table 2. A *t* test of the difference between the means was not significant, nor was an *F* test (*df* 23, 29) of the differences between the variances.

In general, the manic-depressives performed in a manner which would be consistent with their being high value Achievers on all of the tasks administered. On none of the measures did the manic-depressives perform as would be expected of high need Achievers.

Effects of Age and Educational Level

Age and educational level have been reported to affect several of the experimental measures used in this investigation. The effects of these variables were partially controlled by restricting the age range (25–65) and by requiring a sixth-grade reading ability. In addition, the effects of age and education within the range of subject variability were examined. Mean scores on each of the experimental measures for manic-depressives and for controls above their respective median ages were compared with mean scores for those below the median ages; similarly mean scores for *S*s who exceeded the median educational level were compared with scores for *S*s who were below the median educational level (Becker, 1958).

TABLE 2

COMPARISON OF MANIC-DEPRESSIVES AND CONTROLS FOR SUBJECTIVE MOOD STATE

Statistic	Manic Depressive (*N* 24)	Control (*N* 30)	*p*
Mean	24.69	23.82	*ns*
Variance	17.31	11.36	*ns*
Range	19.0–35.5	18.3–30.3	

TABLE 3

COMPARISON OF MANIC-DEPRESSIVES AND CONTROLS ON SIX EXPERIMENTAL MEASURES

Measures	Manic Depressives-(*N* 24)			Controls (*N* 30)			*p*
	Median	x	*SD*	Median	x	*SD*	
v Achievement	4.83	4.86	1.00	4.23	4.16	.67	<.02
F scale	4.62	4.57	.87	4.17	4.09	.83	<.05
TFI scale	4.29	4.19	.73	3.76	3.67	.67	<.01
n Achievement	4.25	5.38	3.98	5.00	5.77	4.68	.88
Xs-in-Circles	562.00	571.17	122.28	560.00	573.83	114.92	.83
Additions[a]	505.00	534.58	230.17	388.50	497.13	220.	.37

[a] Scores are in total seconds to complete the 200 addition problems.

Value Achievement scores increased significantly with age for the manic-depressives ($p = <.05$) but not for the controls. Whereas level of education was inversely related to scores on the F scale ($p = <.05$), and TFI ($p = <.05$) for the controls, no significant relation was obtained for the manic-depressives.

In order to determine whether the controls were representative of nonpsychiatric populations generally, their performance was compared with normative data for comparable age and educational levels. No significant differences were found (Becker, 1958).

DISCUSSION

Manic-depressives in this study performed like high value Achievers with respect to achievement striving, attitudes toward authority, and conventional beliefs. In the manic-depressives as in the high value Achievers, achievement and conformity behavior seem to have the goal of gaining the approval of others rather than of satisfying internalized standards. The defensive conforming nature of the manic-depressive's achievement strivings, and the rigidity of his highly stereotyped, conventional attitudes are further supported by the effects of age and educational level on these variables. Whereas achievement values were independent of age for the control *S*s, the strong conscious valuation of achievement for manic-depressives increased with age. Pros-

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

THE INTERPRETATION OF OPINION STATEMENTS AS A FUNCTION OF RECIPIENT ATTITUDE[1]

MELVIN MANIS

Veterans Administration Hospital, Ann Arbor, Michigan

ALTHOUGH there is general agreement that a person's attitudes influence his interpretation of opinion statements, the direction and magnitude of this influence is as yet unclear. One theory, which has received perhaps its clearest statement from Festinger (1957), leads to the prediction that in interpreting statements of opinion, the recipient minimizes the apparent disparity between his own views and the views of the communicator. Presumably, so doing reduces *dissonance* and the tension that it commonly produces. Message distortion of this type has been reported by a number of investigators (Allport & Postman, 1952; Cooper & Jahoda, 1947; Edwards, 1941; Prothro, 1955).

The views of Hovland, Harvey, and Sherif (1957) and their associates stand in marked contrast to Festinger's (1957) position, although to the best of the writer's knowledge there has been scant recognition of this fact and a critical comparison of the two theories has not been attempted. Instead of focusing attention on the motivational consequences of communicator-recipient disagreement, the Hovland-Sherif group rely heavily on the results obtained in psycho-physical experiments. They cite a number of studies which show that in many judgmental tasks, introducing anchors (standards) which are at the end of the series of stimuli to be judged, or slightly removed from the end, results in a displacement of the scale of judgment *toward* the anchor (the so-called "assimilation effect"). When the reference point is too remote, however, the displacement is *away* from the anchor (the "contrast effect"). Extrapolating from these findings, Hovland, Harvey, and Sherif (1957) are led to the possibility that "when *S*s have established attitudes and are personally involved in a controversial social issue, their 'own stand' functions as the major anchorage affecting reaction to and evaluation of communication. In this case, communication near *S*'s own stand would be assimilated to it, while communication at variance with *S*'s own stand would be displaced still further away (contrast effect). Whether assimilation or contrast effects appear would be a function of the relative distance between *S*'s own stand and the position of communication." These investigators further speculate that contrast effects should begin to appear when the message advocates a position that is unacceptable to the recipient; if the message advocates a position that is within *S*'s "latitude of acceptance," or slightly beyond this range, assimilation effects are predicted.

Prothro's results (1957) are in general accord with these principles, as are the findings of Hovland, Harvey, and Sherif (1957), who originally proposed the theory; in the latter study, however, the data obtained from *S*s who were expected to show the assimilation effect offered only weak support for the theory. Other studies (Hovland & Sherif, 1952; Sherif & Hovland, 1953) have reported contrast effects, without revealing any evidence of assimilation under conditions of minimal conflict between the views of communicator and recipient.

In addition to their work relating attitudes to message interpretation, Sherif and Hovland (1953) have shown that attitudes can affect the extension or constriction of the scale that the individual uses in judging opinion statements. These investigators asked white and Negro *S*s, who were chosen because of their conflicting social attitudes, to sort statements about the Negro into as many categories as they considered necessary; Negro *S*s used fewer categories than did the whites. The results were interpreted as indicating that "intensely involved individuals with an extreme position are more likely to adopt a constricted scale (with few categories)."

---

[1] At the time that this study was conducted, the author held a Public Health Service career-teacher grant at the University of Pittsburgh. The author would like to thank E. Lowell Kelly and Milton Rokeach for their editorial suggestions in the preparation of this manuscript; the efforts of Gloria T. Horn and David Casey, who helped in the analysis of the data, are also gratefully acknowledged.

All statements are those of the author and do not necessarily represent the opinions or policy of the Veterans Administration.

The present experiment represents an attempt to obtain additional data pertaining to the effect of recipient attitude on (*a*) message interpretation and (*b*) scale constriction.

METHOD

Messages

A class of 35 undergraduate students wrote anonymous statements explaining their views on college fraternities. From this pool, four profraternity, four antifraternity, and four relatively neutral messages were selected for inclusion in the experiment. The primary criterion in selecting the messages was the authors' ratings of the concept *college fraternities* on the following evaluative scales: kind–cruel, valuable–worthless, honest–dishonest, fair–unfair, nice–awful, and good–bad. The six scales were presented graphically in a seven-interval format and were arbitrarily rotated to reduce response set, i.e., the "good" pole was on the left for half the scales and on the right for the other half. The obtained ratings were summed across scales, after a value of 1 had been assigned to the "good" end and 7 to the "bad" end of each scale. Previous research by Osgood and his collaborators (Osgood, Suci, & Tannenbaum, 1957) has shown that attitude scales of this type have satisfactory reliability and correlate highly with the more conventional Thurstone and Guttman scales. The authors' ratings were obtained immediately after their messages had been completed.

Recipients

The experimental messages were presented to 70 undergraduate psychology students. Before reading the messages, the recipients' attitudes toward fraternities were assessed by means of the rating scale procedure. They then read the experimental passages and attempted to predict each writer's ratings of college fraternities, again using the six evaluative rating scales; these predictions were summed across scales to yield a measure of each communicator's judged position on the attitude continuum. The passages were assembled in a constant, chance-determined order, with the restriction that no two messages expressing a similar viewpoint could appear in succession.

In order to compare the judgments of *S*s who held varying attitudes, the total sample of recipients was divided into quintiles on the basis of their attitude scores; there were 14 *S*s in each quintile.

RESULTS AND DISCUSSION

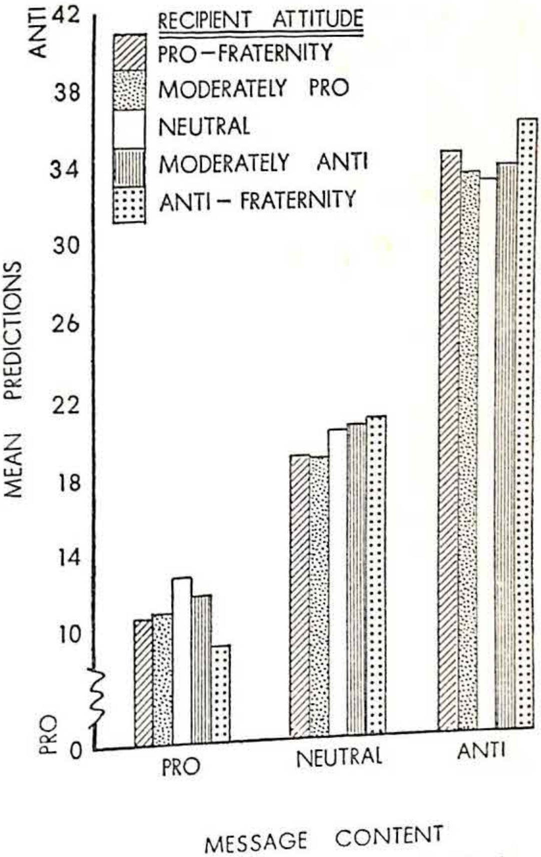


FIG. 1. Mean predictions as a function of attitude.

Figure 1 presents the mean responses of the five recipient-groups to the profraternity, neutral, and antifraternity messages; these responses represent the recipients' predictions as to the views held by the writers. The results for each type of message are shown separately, with the recipient-groups arranged from left to right, in order of decreasing favorability toward fraternities; predictions that are high on the ordinate reflect a belief that the communicator *disliked* fraternities. An analysis of variance revealed that in responding to the antifraternity passages, the recipient groups differed significantly ($p < .05$); for the profraternity and the neutral messages, the group differences fell short of significance ($p < .10$).[2]

Interestingly enough, the responses to the "pro" and the "anti" messages show curvilinear trends that are essentially mirror images of one another. In both cases, the more strongly committed the *S*, the greater his tendency to regard the communicators as holding relatively extreme views. While these results seem at variance with Festinger's position, they can be interpreted in terms of the Hovland, Harvey, and Sherif model, if we assume that each *S*'s latitude of acceptance included at least two types of messages. Specifically, it is assumed (*a*) that a pro *S* would find both pro and neutral messages within his latitude of acceptance, (*b*) that an anti *S* would regard anti and neutral messages as acceptable, and (*c*) that a neutral *S* would regard messages on both sides of his own position, i.e., both pro and antimessages, as being within his latitude of acceptance. Starting with these assumptions, and focusing attention on the extremely pro, the extremely anti, and the neutral *S*s, the data obtained from the profraternity messages can be interpreted as follows: the pro *S*s appear to have displaced the communicators *toward* the end of the scale that they themselves favored (the assimilation effect); that is, the mean prediction of these *S*s was more profraternity than the mean prediction of all *S*s combined. Since profraternity messages were within the pro's latitude of acceptance, a shift of this sort would be consistent with the Hovland, Harvey, and Sherif theory. The anti *S*s seem to have displaced the procommunicators *away* from their own anti stands (the contrast effect),[3] a result which is also consistent with the theory, since profraternity messages would doubtless be outside the anti's latitude of acceptance. Finally, the neutral *S*s seem to have shifted the procommunicators toward the midpoint of the scale (assimilation); this too is in accord with the theory, if we assume (as above) that neutral *S*s would regard pro messages as acceptable. Similar reasoning can be applied to the data obtained from the antifraternity messages.

---

[2] With 4 and 65 degrees of freedom, an F of 2.51 was needed for significance at the .05 level. For the antifraternity messages, F equalled 2.79; for the profraternity messages, F equalled 2.49; for the neutral messages, F equalled 2.29.

Application of the analysis of variance for trends (Grant, 1956; Grandage, 1958) revealed significant quadratic components ($p < .01$) for the profraternity and antifraternity messages and a significant linear component ($p < .01$) for the neutral messages.

The predictions made by the moderately pro- and moderately antifraternity *S*s, in response to the two extreme groups of messages, are somewhat more difficult to interpret. These *S*s apparently responded much as the neutrals did, in that their general tendency was to assimilate both pro- and antifraternity messages toward their own positions. In order to integrate these data with the Hovland-Harvey-Sherif theory, one would have to assume that both extreme viewpoints were within the "moderates'" latitude of acceptance, or slightly beyond this range.

On the neutral messages, there was a linear relationship between attitude and interpretation of messages; *S*s who favored fraternities felt that the neutral communicators were more profraternity than did *S*s who opposed them. While this assimilation trend is clearly in accord with Festinger's viewpoint, it is also consistent with the Hovland, Harvey, and Sherif model, if we assume (as above) that neutrality was an acceptable position to all recipients.

It is interesting to contrast the results of the present experiment with those reported by Segall (1959). This investigator had university students make absolute judgments of statements about fraternities and found virtually no differences between judges of varying attitude. He reasons that his negative results may have occurred because of the relatively narrow range of opinions represented in his *S*s, and this may indeed be a crucial difference between the two studies, for the present data were collected on a big city campus, where the fraternity system was probably more controversial than it was in the setting that Segall describes. It is also possible that the different methods employed in the two studies were crucial in determining their conflicting results; that is, while Segall's *S*s were asked to judge the extremity of opinion statements, *S*s in the present experiment did not judge the statements; instead, they predicted the positions advocated by the persons who had made the statements. It may be that this latter set is more likely to call the *S*'s own views into play as an implicit standard and determinant of his response.

Attitude and Scale Constriction

Scale constriction was measured by computing a standard deviation for each *S*'s predictions over the set of 12 messages. Figure 2 presents the mean standard deviation for each of the five groups of recipients. The groups are again arranged in order of decreasing favor-

---

[3] This statement and subsequent statements regarding judgmental shifts are based on a comparison of the mean prediction given by a specific group of *S*s (here, the antis), and the mean prediction of all *S*s combined.

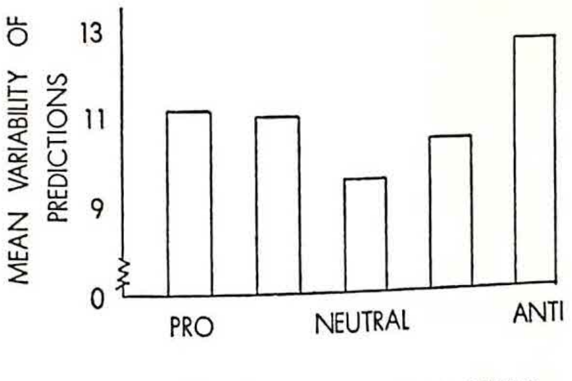


FIG. 2. Variability of predictions as a function of attitude.

ability toward fraternities, with the profraternity *S*s on the left and the antifraternity *S*s on the right. An analysis of variance revealed that the difference between the group means was significant at the .025 level. As will be seen, the relationship between attitude and scale constriction is curvilinear. *S*s with relatively neutral views perceived the messages as reflecting a narrower range of opinion than did *S*s who were committed on the fraternity issue; apparently the neutral *S*s felt that there were only minor differences between the various communicators. This trend is perhaps seen most clearly if Fig. 2 is "folded over," resulting in the formation of three recipient-groups: the highly committed *S*s, the moderate *S*s, and the neutral *S*s. The mean standard deviations for these groups are 11.9, 10.7, and 9.6 respectively.

It is likely that these constriction differences represent a "byproduct" of the displacement trends that were discussed in connection with Fig. 1. That is, the neutral and moderate *S*s may show a more constricted scale of judgment than the committed *S*s because they tended to displace the communicators *toward* the midpoint of the scale, while the committed *S*s displaced the communicators *away* from the midpoint.

Although the results obtained from the pro and the anti messages (see Fig. 1), and the results shown in Fig. 2, could not readily be attributed to random error, they could have been due to differences in response set; more specifically, the relatively uncommitted *S*s may have been unwilling to use the extremes of the scales, regardless of what they were rating. A response set of this sort could have caused these *S*s to (*a*) respond in a neutral fashion to the attitude scales and (*b*) reduce their range of judgments when responding to the experimental messages. To check on this possibility, the five recipient groups were compared in terms of their ratings of two unrelated concepts—jazz and labor unions. These ratings had been obtained in order to familiarize the *S*s with the use of the scales. If the constriction differences discussed above were, in fact, due to differences between the groups in their willingness to use the extremes of the scales, the neutrals' ratings should have shown only slight deviations from the midpoint, the moderates should have shown greater deviations, and the committed *S*s should have been most willing to use the extremes. Examination of the data indicated that this was *not* the case; in fact, the neutral *S*s showed a greater willingness to deviate from the midpoint than did either of the other two groups. The relationship between attitude and scale constriction, which is shown in Fig. 2, is therefore not attributable to generalized response set differences.

While the constricted range of the neutral *S*s' predictions was quite marked in the present study, the obtained results seem somewhat at variance with the interpretation that Sherif and Hovland (1953) offer for their data. As noted above, these investigators found that Negro *S*s used a relatively constricted scale in judging items dealing with the Negro, and concluded that they did so because of their intense involvement with the experimental issue. The results of the present experiment seem to contradict this interpretation, if we make the assumption that the neutral *S*s were probably *less* intensely involved in the fraternity issue than the committed *S*s. While these discrepant findings may be attributable to differences in experimental procedure or topic, it may be that the Sherif and Hovland results were not due to their *S*s' varying degrees of involvement in the experimental task; instead, they may reflect the operation of cultural, educational, or intellectual factors which might have caused the Negro *S*s to use a relatively constricted scale, regardless of the stimulus materials that they were asked to judge.

SUMMARY

Five groups of college students, who varied in their attitude toward fraternities, were presented with the written statements of 12 communicators, who had been asked to explain their views on fraternities. There were three types of communicators: those who favored fraternities, those who opposed them, and those who were relatively neutral. After reading each message, the recipients were asked to indicate the communicator's attitude toward fraternities, as defined by his response to an attitude scale.

In responding to the profraternity and the antifraternity messages, there was a curvilinear relationship between attitude and judgment: in both cases, the committed *S*s perceived the communicators as occupying more extreme positions than did the neutral *S*s. In responding to the neutral messages, the relationship between attitude and judgment was linear: *S*s who favored fraternities felt that the neutral communicators were more profraternity than did the *S*s who opposed them. These results were interpreted as being generally consonant with a theory developed by Hovland, Harvey, and Sherif (1957).

A related finding was the fact that the neutral *S*s used a relatively constricted scale of judgment; they clustered their predictions within narrower limits than did the committed *S*s, apparently feeling that the messages did not reflect a particularly wide range of opinion. If we make the assumption that the neutral *S*s were probably *less* intensely involved in the fraternity issue than the committed *S*s, these results appear to contradict Sherif and Hovland's contention that intensely involved individuals with an extreme position are likely to adopt a constricted scale of judgment.

REFERENCES

ALLPORT, G. W., & POSTMAN, L. J. The basic psychology of rumor. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley, (Eds.), *Readings in social psychology.* (rev. ed.) New York: Holt, 1952. Pp. 160–170.

COOPER, EUNICE, & JAHODA, MARIE. The evasion of propaganda: How prejudiced people respond to anti-prejudice propaganda. *J. Psychol.* 1947, **23,** 15–25.

EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36,** 34–61.

FESTINGER, L. *A theory of cognitive dissonance.* White Plains, N. Y.: Row, Peterson, 1957.

GRANDAGE, A. Orthogonal coefficients for unequal intervals. *Biometrics*, 1958, **14,** 287–289.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psych. Bull.*, 1956, **53,** 141–154.

HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47,** 822–832.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reaction to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55,** 244–252.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PROTHRO, E. T. The effect of strong negative attitudes on the placement of items in a Thurstone scale. *J. soc. Psychol.*, 1955, **41,** 11–17.

PROTHRO, E. T. Personal involvement and item displacement on Thurstone scales. *J. soc. Psychol.*, 1957, **45,** 191–196.

SEGALL, M. H. The effect of attitude and experience on judgments of controversial statements. *J. abnorm. soc. Psychol.*, 1959, **58,** 61–68.

SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48,** 135–141.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55,** 150–155.

(Received February 27, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

COGNITIVE CONSISTENCY AND ATTITUDE CHANGE

WILLIAM J. McGUIRE
University of Illinois

THERE has been a recent resurgence of interest among psychological researchers in the predictive value of a postulate that people tend to maintain logical consistency among their cognitions (and even between cognitions and more gross behavior). The most influential single source of this revived interest is probably Heider's (1946, 1958) "balance" theory, while Festinger's (1957) theory of cognitive dissonance is probably the most stimulating elaboration to date. Most of the approaches that employ this postulate recognize, implicitly or explicitly, that the tendency towards consistency is not absolute. Its expression is limited by other, potentially conflicting forces in the person and by the inadequacies of his cognitive apparatus, even in the absence of conflicting forces, to achieve full consistency. These coordinate hypotheses of cognitive consistency and its predictable limitations have been applied successfully to the prediction of behavior in many areas. The work reported here involves two applications in the area of persuasion.

The first of these applications involves what will be called the "Socratic method" of persuasion. The persuasion technique usually studied in the social sciences involves inducing a change in a person's opinion by presenting him with a persuasive message from an external source, either through a "mass" one-way communication channel or in a face-to-face reciprocal communication network. The postulate of cognitive consistency suggests that persuasion could be effected by the quite different technique of eliciting the persuasive material from the person's own cognitive repertory, rather than presenting it from outside. Specifically, this technique would involve asking the person his opinions on logically related issues, thus sensitizing him to any inconsistencies that exist among his stands on these issues and producing a change towards greater mutual consistency. The technique has an obvious affinity to the "Socratic method" used in the Platonic dialogues (Jowett, 1937).

The derivation of the foregoing hypothesis requires, besides the postulate of a tendency to maintain cognitive consistency, the assumptions that various distorting factors (of which "wishful thinking" is given primary attention in this study) have introduced inconsistency among the initially elicited set of opinions, and that the temporally contiguous elicitation of these opinions brings the tendency towards consistency more strongly to bear upon them.

Part of the study described below was designed to test the hypothesis that a person tends to change his opinions on logically related issues in the direction of greater mutual consistency when he is merely asked to state these opinions in close temporal contiguity. Several additional hypotheses regarding the focus of the Socratic effect, that is, the type of opinion most likely to be adjusted in case of inconsistency, were also tested.

The second application of the cognitive consistency postulate to the area of persuasion in the present experiment dealt with the logical repercussions of persuasive communications. The effectiveness of a persuasive message is usually gauged by the amount of change in opinion that it induces on the issue with which it explicitly deals. From the postulated consistency tendency, however, it would follow that an induced change on the explicit issue tends also to produce changes on logically related (unmentioned) issues such as would tend to maintain internal consistency among beliefs.

From the additional, limiting assumption that there is a certain amount of inertia in the person's cognitive apparatus, two additional hypotheses were derived. It was hypothesized that the consistency-maintaining change expected on remote issues is less than the amount logically required for complete consistency with the induced changes on the explicit issues. Further, it was hypothesized that inertia would also result in temporal effects, such that the impact of the message on the remote issue occurs only gradually, the opinion on the remote issue continuing to change in the logically required direction for some time after receipt of the persuasive message.

Since most of these hypotheses involve predictions of fine graduations of change towards or away from a reference point of precise cognitive consistency, we need a quantitative definition of consistency. Such a definition is provided by a model combining formal logic and probability theory, the relevant aspects of which are summarized below.

Consider three issues so interrelated that a person taking Stands a and b on two of these issues would, if he were to be consistent, be logically required to take Stand c on the third. That is, the three stands are in a syllogistic relationship, with c following as a valid conclusion from the conjunction of a and b. If the person's opinions on these issues are obtained by having him indicate his adherence to each of the Stands, a, b, and c, on a probalistic scale, we can specify that to be completely consistent, these probabilistically scaled opinions must be interrelated in the form

$$p(c) = p(a \,\&\, b) + p(k)(1 - p(a \,\&\, b)) \tag*{[1]}$$

where $p(k)$ is his opinion of the probability of c on bases other than the conjunction of a and b. It can also be specified that if we induce a change of opinion on a and/or b (for example, by a persuasive communication) while $p(k)$ does not change appreciably, then the required change on c necessary for maintaining logical consistency among the opinions is

$$\Delta p(c) = \Delta p(a \,\&\, b) \tag*{[2]}$$

which, assuming that a and b are independent events i.e., that $p\,(a \,\&\, b) = p(a)\;p(b)$ becomes

$$\Delta p(c) = \Delta p(a)p(b) + \Delta\; p(b)p(a) + \Delta p(a)\Delta p(b) \tag*{[3]}$$

where $p(a)$ and $p(b)$ refer to the initial opinions, before the communication induced changes. When, in the discussion below, we refer to the "logically required change on the conclusion," we are referring to the $\Delta p(c)$ value specified in Equation 3.

Method

Material. The *S*s' opinions on sets of logically related issues were measured by means of a questionnaire that contained 48 statements. These statements comprised 16 sets of three syllogistically related propositions; that is, two of the propositions stated stands on a pair of issues, agreement with which would logically imply agreement with the stand taken in the proposition on the third issue.[1] The three propositions making up any one syllogism appeared in different parts of the questionnaire, being interspersed with the propositions of other syllogisms.

The *S*s were asked to indicate in probabilistic terms the extent of their adherence to the stand taken in each proposition. For this purpose each proposition was followed by a graphic scale consisting of a five-inch horizontal line marked off at half-inch intervals and calibrated 0, 10, 20. . . 90, 100. To the left of the scale was written "Very improbable end," and at the right, "Very probable end." The *S*s were asked to decide how probable they felt the truth of each statement to be, and to indicate this subjective probability by drawing a line through the scale at the appropriate point.

In a later section of the questionnaire the propositions were repeated, this time with a scale consisting of five boxes labelled from left to right "very desirable," "somewhat desirable,". . .," "very undesirable." The *S* was instructed to indicate his feeling regarding the desirability of each proposition by putting an X in the appropriate box. (As explained below, these ratings of desirability were obtained to allow identification of initial inconsistencies, due to wishful thinking, among the probability ratings.)

The questionnaires were administered during each of the three experimental sessions. Between the first and second administrations of the opinion questionnaires, a series of persuasive messages were communicated to the *S*s in mimeographed booklets.[2] There were 16 such messages, each directed at one premise of each of the 16 syllogisms. Each message argued that the proposition in question was true, i.e., was aimed at increasing the *S*'s probabalistic estimate of the statement's truth. These messages were "one-sided," 200–300 words long, employed rational appeals, and based arguments on factual evidence. They were mild in tone and attributed to a purportedly competent and dependable source.

---

[1] A complete list of all the syllogisms and persuasive messages used in the present study has been deposited with the American Documentation Institute. Order Document No. 6235, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

One of the 16 syllogisms used was as follows. Students who violate any regulation that has been made to safeguard the lives and property of other students will be expelled; the regulations against smoking in the classrooms and corridors were made to safeguard the lives and property of the student; students who violate the regulation against smoking in the classrooms and corridors will be expelled.

[2] The text of all 16 persuasive messages can be obtained in the ADI Document No. 6235, cited in Footnote 1.

Procedure. There were three 50-min. experimental sessions. In the first, the questionnaires were administered to obtain "before communication" measures of opinions on the issues. The second session, which came one week after the first, was devoted to communicating the persuasive messages and then to a second administration of the opinion questionnaires in order to measure the immediate impact of the persuasive messages. One week later, the *S*s participated in the third session during which the questionnaires were again administered (in order to ascertain the delayed effects of the persuasive messages). At this final session the *S*s were also told the purpose of the experiment and the *E* pointed out the deceits that he had employed in the experiments and his reasons for so doing.

Until this final "catharsis," an attempt was made to conceal, or at least distract attention from, the persuasion aspect of the experiment and from the investigator's interest in cognitive consistency. The study was purported to be an investigation of people's ability to understand experts' discussions of controversial topics. The persuasive messages were alleged to be excerpts from such discussions, and after each message the *S*s were required to answer a series of comprehension questions (to emphasize the contention that we were primarily interested in studying the *S*s' understanding of the material). The opinion questionnaires were explained as required to ascertain whether the issues dealt with were actually controversial in the group being studied and whether they continued to be controversial throughout the period of the experiment.

Subjects and design. A total of 120 *S*s participated in the experiment. One-fourth of these were high school seniors and the remainder, college freshman; the ratio of men to women was about 2:1 at both grade levels. To minimize the likelihood that the consistency hypotheses would receive specious confirmation by the employment of atypically rationalistic *S*s, these *S*s were selected from a group with a rather low level of intellectual attainment. Their modal high school grade average was below the 30th percentile for all high school students in their state. Each of these students reported that he had never had a course in formal logic or in probability theory.

In order that the *S*s might have time to respond thoughtfully to the opinion items and to read the persuasive messages carefully, each *S* was required to provide data on only 8 of the 16 syllogistically related sets of propositions. On only four of these eight did he receive persuasive messages, the other four serving as no-communication "control" items. These no-communication syllogisms provided the data for testing the Socratic effect hypotheses. The syllogisms on which messages were communicated provided the data for testing the hypotheses concerning logical repercussions. Specific syllogisms were alternated between message and no-message conditions from *S* to *S*.

Results and Discussion

Hypotheses Concerning the Socratic Effect

Initial inconsistencies among the opinions. In order to identify initial opinion inconsistencies we employed an auxilliary hypothesis that the subjective probabilities of the propositions would be distorted in the direction of wishful thinking. Internal evidence indicates that this assumption was a valid one. The 16 syllogisms were partitioned into two subsets: the eight in which the desirability of the conclusions was low relative to that of the premises; and the eight whose conclusions had relatively high desirabilities. Wishful thinking would tend to make the subjective probabilities of the conclusions low relative to those of their (more desirable) premises in the first subset of syllogisms, while the subjective probabilities of the conclusions in the second set should be high relative to their less desirable premises. This prediction is borne out by the subjective probability ratings obtained in the first session, as shown in Table 1. (To avoid confusing them with the significance levels, the obtained subjective probability ratings given in this report are the obtained probability ratings, multiplied by 100.) The mean subjective probability of the conclusions exceeded the mean joint probability of the premises by only 2.20 points on the 100-point scale in the first eight syllogisms, while in the other eight syllogisms (the conclusions of which had relatively high desirability) the mean subjective probability of the conclusions exceeded the mean joint probability of the premises by 17.18 points. This difference between 2.20 and 17.18 points is significant beyond the .01 level. (All significance levels reported in this paper are based on analyses of variance and two-tailed tests.) Hence, the attempted manipulation of initial inconsistency by use of the tendency toward wishful thinking was successful.

The main effect. The changes in opinion obtained in the conditions in which no persuasive message was communicated on the given issues provide the test of the main Socratic hypothesis, that when the person's opinions on related issues are elicited in close temporal contiguity, his need for consistency is brought more fully to bear on them. Hence, when these opinions are again elicited on a later occasion, they will be found to have changed towards greater mutual consistency. In view of the initial distortions from consistency produced by wishful thinking, it would follow that in the no-communication conditions in the

TABLE 1

SUBJECT PROBABILITY SCORES (MULTIPLIED BY A FACTOR OF 100) IN THE NO-COMMUNICATION CONDITIONS

	Session	Major Premises	Minor Premises	Joint of Premises	Conclusions	Excess of Conclusions over Premises
Syllogisms with conclusions more desirable than premises	1st	62.15	54.14	33.66	35.86	2.20
	Changes, 1st to 2nd	−3.04	1.45	−0.79	5.17	5.96
Syllogisms with conclusions less desirable than premises	1st	55.50	55.41	30.12	47.30	17.18
	Changes, 1st to 2nd	3.96	−1.98	2.06	0.35	−1.72

present experiment initial questioning should produce an increase in the subjective probabilities of the conclusions relative to those of the premises in the first eight syllogisms (those with the relatively undesirable conclusions); while for the second eight syllogisms, there would be a relative decline in the subjective probabilities of the conclusions.

The results bear out this prediction. As can be seen in Table 1, the gain in probability of the conclusions over that of the premises in the first eight syllogisms averaged 5.96 points. The probabilities of the conclusions of the second eight syllogisms, on the other hand, showed a loss of 1.72 points relative to the premises. This net difference of 7.68 points between means is significant at the .02 level. Hence, simply eliciting opinions on logically related issues does tend to move those opinions toward greater mutual consistency when subsequent elicitation occurs one week later.

The difference obtained between the two sets of syllogisms as regards internal changes is in the direction that would be expected on the basis of a simple regression from the initially obtained difference. In order to test whether there is a Socratic effect over and above simple regression, an additional covariance analysis was carried out. The obtained difference in change of probability scores between the two sets of syllogisms was adjusted for the difference in initial levels by means of the obtained within-set correlation between the initial probability scores and the changes therein. Even when the obtained changes are adjusted for initial levels by this analysis of covariance, the difference in changes remains significant at the .02 level, indicative that a significant Socratic effect remains, even when the effect of regression is partialed out.

The Socratic effect tends to be produced primarily by the first contiguous elicitation of the opinions, with only slight evidence of an accumulative effect from subsequent, additional elicitations. As discussed above, there was a net Socratic effect, from the first to second elicitation of the opinions, of 7.68 points. The corresponding net effect, from second to third elicitations was again in the predicted direction but amounted to only 2.62 points, which falls far short of the conventional levels of statistical significance.

Focus of the Socratic effect. An attempt was made to answer the further question of where this Socratic effect is focused; that is, which of the person's initially inconsistent opinions are most likely to be changed in bringing about the greater consistency? The present data were analyzed to ascertain whether the amount of change was related to any of three variables: the logical status of the propositions, their initial valence, or their ordinal position on the questionnaire in the first session.

As regards logical status of the proposition, we tested whether the major premise, minor, and conclusion contributed differentially to the change towards consistency. All of the syllogisms used in this study were of the "Barbara" type. Hence the major premise contained the M-P terms; the minor, the S-M terms; and the conclusion, the S-P terms (where S refers to the subject of the conclusion; P to its predicate; and M to the middle term).

Logical status of the proposition did have a significant overall impact on amount of opinion change. As can be seen in Table 1, the major premise contributed most of the change towards increased consistency, the mean change being 3.50 points on the 100-point scale. The conclusion showed the next greatest change towards consistency, a mean of 2.41 points; and the minor premise contributed least, in fact their mean change was slightly (1.71 points) away from consistency. These differences among the three means are significant at the .05 level by an overall analysis of variance.

The overall significance of the variance among the three means is due primarily to the greater change towards consistency (averaging 2.95 points) of the propositions with the P terms (namely, the major premises and the conclusions) than of those without these P terms (namely the minor premises), which averaged a 1.71 point change away from consistency. The difference between the two classes of propositions is significant at the .03 level. It appears to be explicable on the basis of initial distortion. In these "Barbara" syllogisms, the P terms tend to have the greatest emotional valence. Therefore, by the auxiliary hypothesis of wishful thinking, the propositions that include these emotionally-tinged terms would have been the most distorted originally. The minor premises, without the P terms, are typically statements of class inclusion ("Socrates is a man") and therefore less likely to be distorted. Hence, the greater change towards consistency of the other propositions appear to be a consequence of the fact that the others had contributed more of the original inconsistency.

Two other propositional characteristics were investigated for possible relationship to the amount of change in opinion produced by the Socratic method: initial valence and initial order of presentation. It was found that somewhat more change towards consistency occurred on the highly valenced propositions—the 12 most and the 12 least desirable propositions—than on the 24 moderately valenced ones. The direction of this outcome tends to reverse the "polarity" prediction (Osgood & Tannenbaum, 1955; Tannenbaum, 1956), but it does not approach the conventional level of statistical significance. Regarding initial order, it was found that somewhat more change toward consistency occurred on those propositions which, in the first session, appeared earlier on the questionnaire than on those that appeared towards the end, but that this difference was not statistically significant. The difference has been predicted on the assumption that the later appearing propositions had already, on the first administration of the questionnaire, been brought into closer agreement with the opinions given on the propositions that had come earlier, so that less change remained to be made.

A number of other hypotheses, not tested in this study, regarding propositional characteristics that affect amount of change in cases of mutual inconsistency could be tested. Opinions on more familiar issues may be more stable than those on less familiar. Opinions on issues that show greater cognitive complexity and greater articulation with the rest of the person's cognitive system (as defined, for example, by Zajonc's [1954] procedures) might participate less in the change towards consistency than more isolated and undifferentiated opinions. Tannenbaum's (1956) finding (that when differently valenced sources and stands on issues are paired, the person's attitude on the issue shows less change than his attitude towards the source) is suggestive in this regard. Following Asch's (1952) thinking that persuasion involves a change of the issue being judged rather than change of judgement on the given issue, we might expect that in case of conflict, the greatest change would occur on the issues most easily redefined. Another hypothesis is that the relative proneness to readjustment of conflicting opinions is related to the bases for accepting the opinions. For example, following Kelman's (1956) terminology, we might conjecture that opinions based on internalization might be more prone to change, in case of internal conflict, than those based on identification or on compliance. These latter need not be consistent with one's other beliefs, because the adherence to them is based on extrinsic considerations.

Hypothesis Concerning Logical Repercussions

The results just reported regarding the Socratic effect hypotheses are all based on data in the no-communications conditions. The following results, which deal with the logical repercussions of persuasive messages are based on data from the conditions in which persuasive communications were received. The changes reported as due to the persuasive messages are computed from a baseline provided by the change from first to second session in the no-communication condition. Since the results concerning the Socratic effect indicate certain systematic changes in opinions occur from first to second session even in the absence of persuasive messages, the message-induced changes must be computed from a control baseline that will eliminate those other changes and also any regression trends.

Immediate postcommunication changes. The induced changes of opinion on the explicit target issues (the premises) were considerable. Each message argued that the stand taken in one of the premises of each syllogism was true and that, therefore, the receiver should increase his subjective probability regarding the truth of that proposition. As can be seen in Table 2, the mean induced increase on the 16 explicit propositions immediately after the communication was 17.56 points on the 100-point scale ($p < .001$).

In order to maintain cognitive consistency, the *S*s should therefore have increased their subjective probabilities regarding the truth of the conclusions. Even though neither conclusions nor the additional premises needed for their derivation had been adverted to in the communications, a considerable immediate postcommunication increase was indeed produced on these conclusions. As can be seen in Table 2, there was an increase (as required for consistency) immediately after the messages on 14 of the 16 unmentioned conclusions, the mean increase being 5.96 points. A mean change of this magnitude is significant at the .01 level. Hence, the results do indicate a consistency tendency sufficiently strong to cause a persuasive message to have an impact on logically related issues not explicitly mentioned in the communication.

These obtained changes on the derivative issues are, however, less than the amount required for complete logical consistency as determined by substituting the amount of the obtained change on the explicit premise in Equation 3. The mean required increase on the conclusions so computed was 11.57 points, while the actually obtained increase was only 5.96 (see Table 2), a difference significant beyond the .05 level. Hence, persuasive messages, while they do produce changes on remote, unmentioned issues, produce less than the amount logically required, a finding that supports the hypothesis of a certain amount of cognitive inertia.

Delayed postcommunication changes. A considerable amount of the immediately induced persuasion on the explicit target issues had dissipated by the end of the week following the communication. As can be seen in Table 2, almost half of the immediately induced change on these explicit issues had dissipated one week later. The mean loss of 8.43 points was significant at the .02 level. On the other hand, the 9.13 points of change remaining after one week was also significant (at the .01 level), indicating that there is still a significant amount of change of opinion on the explicit issue one week after the persuasive communication.

When this loss on the premises is substituted in Equation 3 the amount of the originally induced change on the unmentioned derived issues that should logically have dissipated during the succeeding week was computed as 6.69 points (see Table 2.) The obtained loss during the interim, however, was only a negligible 1.48 points, a gain of 5.20 points relative to the explicit premises for the week ($p = .10$). In fact, by the week after the persuasive communication, the net mean change on the conclusions from their precommunication level was an increase of 4.47 points, which was almost as much as the 4.89 points logically required in view of the net changes on the premises. The trivial extent to which the change on conclusions fell short of the required amount one week after, as compared with the significant extent to which it fell short immediately after the communication, gives some support to the hypothesis that the persuasive messages had a delayed-action effect on the derived issues. The net change on the derived issues did not actually increase with the passage of time, but it fell off less sharply than would have been expected on the basis of the usual rate of temporal decay of the immediately transmitted opinion change, suggesting that continued seepage of the impact on the explicit issue almost completely overcame the opposing decay effect on the derived, unmentioned issue.

Comparison with previous studies. One methodological artifact in the present study, the use of a before-after design, may have augmented the amount of the logical repercussions of the persuasive messages. In order to obtain each *S*'s precommunication opinions he was required to fill out, one week prior to receipt of the persuasive messages, an opinion questionnaire containing all the issues, including the auxilliary premises and conclusions not mentioned in the message itself. It is possible that having stated his opinions on these related issues during the previous week may have somewhat increased their saliency, so that they were more affected by the message than would have been the case had they not been raised the previous week.

On the other hand, the finding of logical repercussions in the present experiment is particularly convincing in view of the fact that the method included a number of conditions that have been shown in previous studies to militate against logical repercussions. In the first place, not only was the conclusion itself unmentioned in the messages but also one of the two premises necessary for its derivation was left unmentioned. Also, the *S*s employed were of a type, being below average in intelligence, that has been found (Cooper & Dinerman, 1951) to be relatively resistant to nonexplicit messages. Furthermore, no mention was made during the experiment of our interest in consistency, and the logically related propositions were widely interspersed with other propositions within the questionnaire. Finally the *S*s gave their opinions on a scale with whose computational axioms they were unfamiliar, so that they could hardly have determined the extent of their inconsistencies even had they been interested in this matter.

Hence, the negative indications reported in some previous studies for the hypothesis that persuasive communications have logical repercussions require comment. While more persuasion has typically been found on a derived issue if that issue is explicitly mentioned rather than only implied, the data reported typically do show some change on the nonexplicit issue (Hovland, Lumsdaine, & Sheffield, 1949; Hovland & Mandell, 1952). Also, the relative lack of effect on the derived issue is most evident in the immediate postcommunication periods. Hovland, Lumsdaine, and Sheffield (1949, pp. 70–71) found that the "Why We Fight" films did have an impact on more general attitudes after the passage of a greater amount of time, and have postulated that there may be need for a "sinking in" period before these derivative effects are felt. Similarly, Katz, Stotland, Patchen, and Jochen (1957) report that giving *S*s case history material regarding the roots of prejudice causes little immediate effect on their own prejudice, but after the passage of several weeks, the extrapolation of the intellectual knowledge to their opinions does occur. Furthermore, the magnitude of the logically required change on the derived issue is, as shown by Equation 3, less than that on the explicit issue. This logically consistent attenuation of the persuasive impact on derived issues due to a compounding of contingencies becomes more severe as the derived issue becomes more tenuously related to the explicit issue. For example, the more premises beside the explicit one needed for its derivation, the less the logically required change on the derived conclusion. Where the derived issues are as remote from those explicitly dealt with as was the case, for example, in the Hovland, Lumsdaine, and Sheffield "Why We Fight" studies (1949), very little change on the derived issues would be logically required even in a perfectly consistent *S*.

However, while only a small change in opinion is logically required on the derived issue, the amount actually occurring is even less. People's opinions seem to follow a "loose-link" model such that a considerable change

TABLE 2

CHANGES IN SUBJECTIVE PROBABILITIES ON EXPLICIT ISSUE AND ON DERIVED, UNMENTIONED ISSUES IMMEDIATELY AFTER AND ONE WEEK AFTER PERSUASIVE MESSAGES

	Obtained Change on Premises	Logically Required Change on Conclusions	Obtained Change on Conclusions	Discrepancies of Obtained from Required Change on Conclusions
Changes from before to immediately after	17.56	11.57	5.96	−5.61
Changes from immediately to one week after	−8.43	−6.69	−1.48	5.20

can be induced on one cognition before any pull is exerted on related cognitions. This effect, however, is at least partially corrected by the continued seepage of change from the explicit issue to the derived one as time passes after the receipt of the communication.

The question of conscious awareness arises in connection with the psychodynamics of the present effect: to what extent is it necessary that the person be consciously aware of the inconsistencies among his beliefs in order that the effects occur? In the present study several procedures were employed in an attempt to minimize the degree of conscious awareness. (For example, no mention was made to the *S*s that their consistency was being assessed; the related propositions were widely separated from one another in the questionnaire; a quantitative expression of opinion was required and the assessment of inconsistency involved a mathematical model unfamiliar to the *S*s.) Hence, the present findings indicate that the effects occur even under conditions that attempt to avoid any suggestion of inconsistency.

The *S*s' level of awareness of inconsistency was not systematically varied within the present experiment, but it is interesting to speculate on the results of such a manipulation. Increasing the person's awareness of his inconsistency by, say, explicitly pointing it out to him might produce greater changes towards consistency. On the other hand, such a procedure could well have the opposite effect by provoking the person's hostility or defensiveness, and so result, not in a greater change towards consistency, but in some other reaction such as the rejection of consistency, with Carlyle, as the hob-goblin of little minds; or in the repression of some of the conflicting issues.

SUMMARY

Two sets of hypotheses were derived from the postulate that people tend to maintain consistency among their opinions on logically related issues. First, it was hypothesized that opinion change would be effected by the Socratic method of simply asking the person to state contiguously his opinions on logically related issues, thus bringing the tendency towards consistency to bear on any logical discrepancies among these opinions. Secondly, it was hypothesized that a persuasive communication directed at a person's opinion on some explicit issue tends to change also his opinions on logically related (even if unmentioned) derivative issues in a consistent direction. A procedure for quantitatively measuring the degree of logical consistency among opinions on related issues was described.

The *S*s were 120 college students participating in three experimental sessions. In the first, before-communication opinions regarding the subjective probabilities and desirabilities of 16 sets of syllogistically related propositions were obtained. In the second session, 16 messages, each arguing for the truth of one premise of each syllogism, were communicated, and the immediate postcommunication opinion of the truth of each of the 48 propositions was obtained. Each *S* served in communication conditions with respect to some issues, and control, no-communication conditions with respect to others. In the third session, one week later, the delayed postcommunication opinions on the truth of the 48 propositions were obtained.

With respect to the Socratic effect hypothesis a significant amount of inconsistency attributable to wishful thinking was found, as postulated on the first elicitation of the opinions ($p < .01$). One week later, on a second elicitation of the opinions, there was a significant ($p = .02$) decline in the amount of this wishful thinking, indicating that the Socratic method did indeed produce the predicted change towards consistency. Several propositional variables were examined for possible relationship to the amount of change towards consistency. It was found that propositions with the predicate term (namely, major premises and conclusions) contributed significantly more ($p = .03$) of the change towards consistency than those without the predicate term.

With respect to the hypothesis concerning logical repercussions, it was found that immediately after the communication there was an induced opinion gain not only on the explicit target issues ($p < .001$) but also the derived issues ($p = .02$), even though neither the derived issues themselves nor even the auxiliary premises needed for their derivation were mentioned in the communication. This

indirectly induced gain on the derived issues fell short, however, of the full amount needed for complete logical consistency ($p < .05$). One week later, there was some suggestion that the impact of the persuasive message had continued to seep down to the derived issues: immediately after the messages the opinions on the derived issues had shown only 52% of the logically required amount of change, while by one week after, this impact had risen to 91% of the required amount ($p = .10$).

REFERENCES

Asch, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

Bagley, W. C. *The educative process*. N. Y.: Macmillan, 1906.

Cooper, Eunice, & Dinerman, Helen. Analysis of the film "Don't be a Sucker": A study in communication. *Publ. opin. Quart.*, 1951, **15**, 243–264.

Festinger, L. *A theory of cognitive dissonance*. Evanston: Roe, Peterson, 1957.

Heider, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.

Hovland, C. I., & Mandell, W. An experimental comparison of conclusion—drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.*, 1952, **47**, 581–588.

Jowett, B. (Ed.) *The dialogues of Plato*. New York: Random House, 1937.

Katz, D., Stotland, E., Patchen, M., & Jochen J. Research in methods for changing attitudes. I. Situational and personality determinants of the reduction of prejudice through self-insight. Univer. of Michigan, 1957. (Mimeo.)

Kelman, H. C. Three processes of acceptance of social influence: Compliance, identification and internalization. *Amer. Psychol.*, 1956, **11**, 261. (Abstract)

Osgood, C., & Tannenbaum, P. The prinicple of congruity and the production of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

Tannenbaum, P. Initial attitude towards source and concept as factors in attitude change through communication. *Publ. opin. Quart.*, 1956, **20**, 413–425.

Zajonc, R. B. Cognitive structure and cognitive tuning. Unpublished doctoral dissertation, Univer. Michigan, 1954.

(Received March 12, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

DIRECT AND INDIRECT PERSUASIVE EFFECTS OF DISSONANCE-PRODUCING MESSAGES

WILLIAM J. McGUIRE

University of Illinois

THE present study was designed to test some applications of Festinger's dissonance theory (1957) to the area of persuasion. Festinger has predicted (1957, Ch. 9) that a person tends to resist a persuasive message to the extent that it tends to increase dissonance. The present study tests a somewhat broader hypothesis. When a person is presented with a message that argues for a position on an explicit issue which tends to bring his opinion on that issue into conflict with other, unmentioned but logically related issues, he can avoid increased dissonance in two ways: (*a*) by resisting change on the explicit issue and (*b*) by changing also on the related (unmentioned) issues. Hence, it is predicted that a dissonance-increasing message, as compared with a dissonance-reducing one, produces (*a*) less change on the explicit issue, and (*b*) more change on unmentioned but logically related ones.

Furthermore, on the assumption that changes induced in a consistency-increasing direction are stabilized by the bolstering effect of the consonant cognitions, it is hypothesized that dissonance-reducing changes persist longer after the presentation of the communication. Hence, it is predicted that both of the consonant types of change—namely, the changes on the explicit issues produced by the consistency-increasing messages and those on the related, unmentioned issues produced by the consistency-decreasing messages— show slower rates of temporal decay than do the corresponding dissonant changes. Inconsistencies were identified on the basis of an assumed tendency toward wishful thinking.

METHOD

Procedure. There were three experimental sessions held at one week intervals during regularly scheduled classes of a freshman English course. Until the post-experimental "catharsis," the study was represented as an investigation of the relation between reading comprehension ability and the readers' beliefs regarding the material read. At each session a questionnaire was administered to measure the *S*s' opinions on the truth of 24 critical propositions. At the first session the use of the opinion scales for the truth judgments and also for measuring the desirabilities of the propositions was explained.

At the beginning of the second session the persuasive messages were communicated and memory questions were asked about the content (to support the fiction that we were studying reading comprehension), and then the opinion questionnaire was administered to determine the immediate effects of the communications. In the third session, the questionnaire was again administered (to measure the persistence of the induced opinion changes) and the true nature of the experiment was then explained to the *S*s.

Material. The opinion questionnaire consisted of a list of 28 propositions. Of these, 24 constituted 8 syllogisms[1] of the "Barbara" type (the other four were repetitions of propositions to provide a reliability check). Beneath each proposition was a scale, described elsewhere (McGuire, 1960), on which the *S* was instructed to indicate his subjective probability regarding the statement's truth. Opinions regarding the desirability of each proposition were obtained on a five-point desirabiltiy scale described by McGuire (1960).

Each persuasive message contained a reasoned appeal, about 250 words long, attributed to a (fictitious) person with a prestige affiliation, arguing in favor of the truth of the premise of one of the eight syllogisms.

Subjects and design. Each of the 35 college freshmen participating in this study received persuasive messages on four of the issues and served as a no-communication control on the other four issues.

RESULTS

Identification of the initial inconsistencies

In order to test the present hypotheses, it was necessary to determine for each message whether it was arguing in direction that increased inconsistency or decreased it. Since each message was aimed at raising the rated truth of a premise, this determination required knowing whether the initial truth rating of the premise had been inconsistently high or inconsistently low on the first administration. To this end, we postulated a tendency toward wishful thinking, amply demonstrated in previous studies by Cantril (1938) and McGregor (1938), such that the rated truth values of propositions are distorted in the direction of their rated desirabilities. The eight syllogisms were therefore divided into four, the premises of which were rated highly desirable relative to the conclusions and four with premises rated relatively low in desirability. In the case of the first four, the persuasive messages were interpreted as arguing in an inconsistency-increasing direction, and for the second four, in a relatively inconsistency-reducing direction.

---

[1] For a detailed account of the syllogisms and the persuasive messages used in this study, see the ADI document No. 6235 referred to in McGuire (1960) Fn. 1. An example of a syllogism whose premises are rated less desirable than the conclusions (Syllogism 8 in Table 1) is the following: Even should there be a war with Russia, any American city hard to reach by air from the European continent will escape nuclear bomb attack. Our city is hard to reach by air from the European continent. Our city will escape nuclear bomb attack even should there be a war with Russia.

The data from the first session tend to confirm this auxiliary hypothesis concerning wishful thinking. If the person were being perfectly consistent, his subjective probability of the conclusion should have been directly proportional to the product of the subjective probabilities of the premises (assuming that internal consistency among subjective probabilities requires their following the axioms of probability theory, and that the proportional contribution of the premises to the total probability of the conclusions remained constant from syllogism to syllogism). To the extent that, on the contrary, his beliefs were being distorted by wishful thinking, the excess probability of the conclusions, over the product of the probabilities of the premises, should have been greater in Syllogisms 5 to 8, than in Syllogisms 1 to 4. The data (see Table 1) give evidence of a considerable amount of such distortion. For Syllogisms 1 to 4, the mean excess probability of the conclusions was only 8.99, while for Syllogisms 4 to 8 it was 18.90 (throughout this report, the subjective probability scores cited are the obtained probability ratings multiplied by 100, in order to avoid confusion with the statistical significance levels). The obtained difference of 9.91 is significant at the .05 level. (All significance levels in this report are based on two-tailed analyses of variance tests.)

Wishful and logical thinking

While it was not a primary purpose of the present experiment to compare these two tendencies, the data make possible an estimate of the extent to which wishful thinking and logical thinking each affects opinions regarding the truth of the conclusions. If a person's opinions are to be logically related, then his rating of the probability of the conclusion ($p(c)$) should equal the joint rated probability of the premises ($p(a)p(b)$) plus the rated probability of the conclusion on the basis of all circumstances other than those specified in the premises. Therefore, if we assume that these other circumstances contribute about equal proportions of the total probability of the conclusion from syllogism to syllogism, then consistency requires that the probability of a conclusion be proportional to the joint probability of the premises from which it follows. The group data indicate that, by this criterion there is a high degree of consistency: the correlation between the initial $p(c)$ and $p(a) \cdot p(b)$ for the eight syllogisms is .74.

Wishful thinking accounts for a large part

TABLE 1

TRUTH RATINGS ON THE PREMISES AND THE CONCLUSIONS BEFORE AND AFTER MESSAGES WHICH ARGUED FOR THE INCREASED TRUTH OF THE PREMISES AND LEFT THE CONCLUSIONS UNMENTIONED

Syllogism with Relatively Desirable Premises						Syllogism with Relatively Undesirable Premises					
Syllogism	Excess prob. concl. before	Change immediately after		Further change one week after		Syllogism	Excess prob. concl. before	Change immediately after		Further change one week after	
		Premise	Concl.	Premise	Concl.			Premise	Concl.	Premise	Concl.
1	14.99	18.56	14.63	−12.97	−23.38	5	10.07	18.56	9.64	−13.86	1.81
2	16.14	11.07	4.29	−4.29	−4.64	6	14.99	27.13	12.14	−9.58	−11.66
3	11.00	7.14	12.14	−3.15	2.79	7	20.85	26.06	0.35	−14.99	−5.35
4	−6.28	20.71	13.56	−10.00	−1.78	8	29.49	22.85	−15.35	−18.92	−1.07
Mean	8.99	14.37	11.15	−7.60	−6.75	Mean	18.90	23.65	1.70	−14.34	−4.07

Note.—Scores are given in terms of subjective probability of truth multiplied by 100.

of the remaining variance among the conclusions. When the effects of the desirability ratings of the premises and conclusions are removed by partial correlation, the correlation between $p(c)$ and $p(a)p(b)$ becomes .96, which is significant at the .01 level despite the small number of degrees of freedom.

Since these and the previously reported correlation coefficients are computed from group means it is not legitimate to generalize to the individual case. However, even on the basis of group data it is interesting that 90% of the variance among the mean opinions on the conclusions can be accounted for with just two variables.

Effects Produced by the Persuasive Messages

Opinion changes on the explicit issues. The first experimental hypothesis stated that the persuasive messages would produce opinion change on the explicit issues to the extent they argued in a direction that would increase consistency. Since the auxiliary hypothesis concerning wishful thinking was confirmed, it would be predicted on the basis of the first hypothesis that a greater increase in the $p(a)$ opinions should be produced by the messages directed at the relatively undesirable premises, than by those directed at the relatively desirable.

This hypothesis is supported by the finding that the mean increase in probability on the desirable target premises (of Syllogisms 1 to 4 in Table 1) was 14.37, while the mean for the undesirable premises (5 to 8 in Table 1) was 23.65, a difference significant at the .01 level. That this greater impact on the undesirable premises is not a regression artifact can be demonstrated by further analyses. In the first place, there is not a great deal of difference between the initial probability levels of the two sets of premises. The mean probability rating of the relatively undesirable premises is 50, and of the desirable, 55, which is too small an initial difference to account for the 10-point differential in effect by statistical regression. A still more convincing argument against alternative interpretations is provided by considering the changes from a baseline provided by the control data, that is, changes from before to after the communication for those subjects not receiving a communication on the given issue. From this baseline, the differential is even greater: the mean increase induced by the message on the relatively undesirable premises was 30 and on the desirable, 11. This difference again is significant beyond the .01 level.

Effects of the persuasive messages on the related, unmentioned issues. While some studies report finding that persuasive communications produce little if any effect on related but unmentioned issues (Cooper & Dinerman, 1951; Hovland & Mandell, 1952), McGuire (1960) has demonstrated that remote repercussions in the person's cognitive arena do occur, though usually to a lesser extent than would be required for complete consistency with the induced change on the explicit issue. The present hypothesis states that such remote impact on unmentioned issues is greater for messages arguing in an inconsistency increasing direction, since otherwise the already dissonant opinions on these derived issues would get still further out of line with opinions on the explicit issues. Since all the messages in this study were designed to increase the rated truth of a premise, the prediction follows that the increase of the truth ratings of the conclusion produced by the messages aimed at the desirable premises (Syllogisms 1 to 4) should be greater than those aimed at the undesirable premises (Syllogisms 5 to 8). The prediction is supported by the data. The four conclusions of the desirable premises show a mean increase of 11.15 points after the messages, while the conclusions of the undesirable premises increase by a mean of only 1.70.

This finding is especially striking since the changes induced on the desirable premises themselves were, as has just been described, less than the changes on the undesirable premises. Assuming the subjective probability ratings behave according to the axioms of probability theory, we can compute by substituting the induced change on the premises in the following equation:

$$\Delta P(c) = P(a)\cdot\Delta P(b) + P(b)\cdot\Delta P(a) + \Delta P(a)\cdot\Delta P(b)$$

where $P(a)$ and $P(b)$ are the initial opinions on the premises. Substituting the obtained changes on the premises in this equation, it can be computed that, to be consistent, there should have been a mean change of 8.30 for the conclusions of the desirable premises and 17.48 for the conclusions of the undesirable. Hence, the obtained changes on the conclusions of the desirable premises exceeded the required amount by 2.85 whereas the obtained change on the conclusions of the undesirable premises fell short of the required amount by 15.78 points, a difference significant at the .01 level. The differential is reduced when, to control for possible regression effects, the changes in the communication conditions are computed from a baseline provided by the control (no communication) conditions, but the reduced difference is still significant at the .05 level.

The relative persistence of consistent and inconsistent opinion changes. On the assumption that the bolstering effect of consonant cognitions would stabilize changes in a direction reducing inconsistency, it was hypothesized that such changes would be more persistent than the dissonance-increasing changes. Hence, we predicted that as regards the changes on the premises themselves, increases on the undesirable premises would be more lasting than those on the desirable; regarding the derived changes on the conclusions, we predicted, conversely, that increases on the conclusions of the desirable premises would be more lasting than the increases on the conclusions of the undesirable premises.

Neither of these predictions is supported by the data. In regard to the persistence of the induced increase on the explicit issues, 7.60 points of the originally induced 14.37 on the desirable premises dissipated during the week; while 14.34 points of the originally induced 23.65 on the undesirable premises dissipated (see Table 1). Hence, in terms of the arithmetic loss during the week, the dissonance-reducing changes show 6.74 points more decay than those which increased dissonance, reversing the prediction that the dissonance-increasing changes would show more decay. In terms of percentage loss of originally induced increase, the two conditions show approximately equal losses at the end of one week. The prediction regarding greater stability of dissonance reducing changes on the remote issues likewise is not confirmed by the data. The relatively desirable conclusions show a 4.07 point decline during the postcommunication week, and the undesirable conclusions, a decline of 6.75 points. Hence, the dissonance-increasing changes are slightly more persistent than the dissonance reducing ones when measured in terms of amount of decay; and the dissonance-reducing changes slightly more persistent when measured in terms of rate of decay.

DISCUSSION

The confirmation of the hypothesis of greater immediate impact of the dissonance-reducing communications is particularly striking because, in the case of both the explicit and the related issues, the tendency toward wishful thinking works against it. The implication of our finding is that a person can more easily be argued into an increased acceptance of the truth of an undesirable state of affairs than a desirable one. The underlying reason for this, suggested by dissonance theory, is that before the persuasive communication, he has already overrated the truth of the desirable outcome and because of his need for consistency, therefore resists further uprating that would make his expectation become even more dissonant with his cognitions of the available evidence.

The confirmation of both derivations from dissonance theory regarding the immediate impact of the persuasive messages but the complete nonsupport for the derivations regarding the persistence of these immediate effects suggests that predictive power of the theory could be improved if a postulate concerning "salience" were added to it. Such a postulate would recognize that all the person's cognitions relevant to a given opinion might not be equally brought to bear on the opinion in a given situation. Rather, one or another subarea within this total cognitive arena might be made more salient by experimental manipulation. It might often happen that the person's cognition on a given issue could be made to appear dissonant in either one direction or the opposite, depending on which of these subsets of relevant cognitions is made salient.

Applying this formulation to the present study which investigates the person's expectations regarding desirable and undesirable states of affair, we can detect two cognitive forces: wishful thinking and logical thinking. The person presumably had initially arrived at a steady state of expectation on the issues which represented a compromise between the two needs—wish fulfillment and reality orientation—a compromise that was fully consonant with neither. In this study the experimental manipulation involved presenting messages that contained rational arguments regarding the probability of the events in question. We can reasonably interpret this procedure as producing increased saliency of the cognitive realm involving logical thinking, reality orientation and objective evidence. Hence, any underlying dissonance between expectations and evidence would be activated, with a resulting readjustment of the cognitions towards greater consonance with the evidence and less with desires.

After the person leaves the experimental situation, however, the increased saliency of the evidential subarea of his cognitive arena should tend to subside to its pre-experimental level. Consequently, the need for consonance between expectations and desires once again becomes relatively more salient, and the dissonance between these beliefs and hopes is now felt behaviorally to a greater extent. By adding such a postulate of salience, dissonance theory could better fit the present results, both as regards initial impact and persistence of effects.

SUMMARY

Several hypotheses regarding the effectiveness of persuasive messages were derived from Festinger's "Theory of Cognitive Dissonance." It was predicted that messages would be effective in changing opinions on the explicit issues to the extent that they argued in dissonance-reducing directions; and on related, unmentioned issues, to the extent that they argued in a dissonance-increasing direction. It was further predicted that the dissonance-reducing changes would show a slower rate of temporal decay after the communication.

These hypotheses were tested by asking 35 college students their opinions regarding the truth of several syllogistic sets of premises and their valid conclusions. The *S*s were then presented with messages which argued for the increased truth of the premises (leaving the conclusions unmentioned). Immediately after the persuasive messages, and again one week later, the opinions on the truths of both premises and conclusions were again measured.

On the basis of desirability ratings of the propositions and of an assumption concerning wishful thinking, the syllogisms were partitioned into two subsets: those whose premises received inconsistently high truth ratings initially and those receiving relatively low initial truth ratings.

Both predictions regarding the immediate impact of the messages were confirmed. On the explicit issues (the premises), more opinion change was produced by the inconsistency-reducing messages ($p < .01$); on the related, unmentioned issues more change was produced by the inconsistency-increasing messages ($p < .05$). The prediction of greater permanence for the dissonance-reducing changes was not confirmed: the dissonance-reducing and dissonance-increasing changes both showed about the same rate of temporal decay. On the basis of these findings, it was suggested that a postulate concerning "salience" be added to dissonance theory.

REFERENCES

CANTRIL, H. The prediction of social events. *J. abnorm. soc. Psychol.* 1938, **33**, 364–389.

COOPER, EUNICE, & DINERMAN, HELEN. Examination of the film, "Don't be a Sucker": A study in communication. *Publ. opin. Quart.*, 1951, **15**, 243–264.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Roe, Peterson, 1957.

HOVLAND, C. I., & MANDELL, W. An experimental comparison of conclusion drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.* 1952, **47**, 581–588.

MCGREGOR, D. The major determinants of the prediction of social events. *J. abnorm. soc. Psychol.*, 1938, **33**, 179–204.

MCGUIRE, W. J. Cognitive consistency and attitude change *J. abnorm. soc. Psychol.* 1960, **60**, 345–353.

(Received March 9, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SOME CONSEQUENCES OF EXPERIMENTAL REJECTION

ALBERT PEPITONE AND CHESTER WILPIZESKI[1]

University of Pennsylvania

Rejection by others has been cited as the determinant of a large number of different responses, pathological and normal. It is widely believed, for example, that social hostility as reflected in delinquent behavior stems from a parental environment in which the child is somehow isolated and unloved (cf. Bandura & Walters, 1958 for a recent variant of this view). Rejection has also been considered an important basis of suicide and other self-punitive responses, as well as of social withdrawal and apathy (cf. Wolberg, 1944). Finally, a laboratory study by Dittes and Kelley (1956) showed that subjects (*S*s) who were given to believe they had low sociometric status in the group conformed more to the opinions of that group than those who thought they were highly accepted.

It is not clear from these studies whether such diverse responses occur under exactly the same conditions of rejection, and if so, whether they are displayed by the same person. The unanswered question is: what responses can be expected from a given kind and degree of rejection, and theoretically, why?

A number of experiments (Pepitone, 1958) have shown that the magnitude and the negative effects of various "ego-threats" can be reduced when the person perceives certain causal properties relating to responsibility, intentionality, etc. in the threatening situation. On the basis of these results, a theoretical process may be proposed to account for some of the responses reported to be associated with rejection. The assumption is made that rejection from an important group generates more or less severe personal "devaluation anxiety" in most individuals, and that more or less strenuous attempts are made to reduce such anxiety. Anxiety-motivated instrumental behavior may be primarily directed toward changing the social environment. For example, rejected persons may comply or identify with the rejector's opinions, become compulsively friendly, withdraw from the group, or display various other *socially adaptive* responses. However, when socially adaptive responses are unsuccessful, impossible, or when for any reason, the anxiety is irreversible, the person then attempts to make the anxiety more tolerable to himself. In other words, when the rejection or other basis of devaluation anxiety is a *fait accompli*, persons tend to respond in a way that is primarily *self-adaptive* in function.

The basic assumption of the present experiment, suggested by the studies referred to above, is that the person can make irreversible devaluation anxiety more tolerable by explaining or fixing responsibility for it. Specifically, it is hypothesized that "causal attibution" to the two general sources of responsibility is reflected empirically in two kinds of responses:

1. *Self-depreciation*. An "internal" attribution of responsibility for the rejection is shown by some form of self-depreciation on the part of the rejectee. Such a response in effect acknowledges or "accommodates" the personal devaluation implicit or explicit in the rejection and thereby makes it more bearable.

2. *Hostility*. Various forms of aversion toward the social environment may be regarded conceptually as an external attribution of responsibility for the rejection. Thus, by disliking the rejectors, by finding fault with them and the social situation generally, the person "accounts" for the rejection, and palliates the devaluation anxiety generated by it.

An important question is whether the two hypothesized responses to rejection tend to occur in the same person. From one theoretical standpoint, it may be supposed that if in the past the person has succeeded in lowering rejection anxiety by internal attribution (self-depreciation), he should be less likely to produce an external attribution (hostility). That is, the two responses should correlate negatively. On the other hand, causal attribution in general may have been "reinforcing," in which case all responses in that class would tend to be positively correlated in individuals. From a different theoretical slant, a positive correlation would result if the relative contributions to the rejection are uncertain. Where the person is not sure of the specific basis for being rejected, he might be expected to implicate both himself and the rejectors. Thus, an unequivocal prediction regarding the relationship among responses to rejection is not possible.

---

[1] This experiment is one of a series supported by the Office of Naval Research through a contract with the senior author N-ONR 551-(27). We gratefully acknowledge this support, although we must take full responsibility for the views expressed in this paper. We have also benefited from a critical reading of the manuscript by J. Aronfreed and S. Feshbach. We wish to acknowledge the help of R. C. Mackowiak in parts of the statistical analysis.

There are of course, all degrees of rejection. On the assumption that stronger rejection would tend to produce more vigorous self-adaptive responses, an attempt was made operationally to establish Implicit and Explicit levels of rejection. As it turned out empirically, the differences in effects between the two conditions were negligible. *S*s under the two conditions were, therefore, combined for purposes of analysis.

METHOD

Ninety-five college undergraduate male volunteers served as *S*s. Each three-person group consisted of one *S* and two confederates.[2] The behavior of the confederates was completely prescribed in advance by a well rehearsed script.

After assembling the group in the laboratory, *E* informed the *S*s that they were going to take part in an important survey concerned with the breakdown of family structure. The purpose of the study was to investigate the attitudes of college undergraduates toward two issues relevant to family disorganization: (*a*) premarital sex relations, and (*b*) drinking among minors. Specifically, *S*s were asked to reflect upon the question: "Do you, in general, encourage or discourage premarital sex relations?" and to jot down the three strongest reasons for their opinion. Lots were then ostensibly drawn to determine who would discuss his views first. However, the drawing was so contrived that *S* always led off. He declared himself either pro or con on the question and then elaborated his views by citing the reasons he held to be decisive. Finally he was requested to raise up in full view one of a series of cards numbered 1 to 6, representing the degree to which he personally either encouraged or discouraged premarital sex. The identical procedure was followed by the two confederates. In the *No Rejection* condition, the confederates held the same opinions as *S*, and assigned approximately the same numerical weight to them as did *S*. In both the *Implicit* and *Explicit* conditions, the confederates differed from *S* by adopting the diametrically opposed opinion. The reasons given by the collaborators and the numerical strength of their opinions were such as to maximize the isolation of *S* while maintaining the naturalness of the situation.

Following the 15-min. discussion, the groups under the rejection conditions were given a brief recess before taking up the second topic. The *Explicit Rejection* procedure took place during this recess. *E* left the room mentioning that he had forgotten some forms. The two confederates then became engaged in friendly conversation. They were instructed to ignore or to respond curtly to any of *S*'s questions or remarks, and to present an unfriendly demeanor when glancing at *S*. A number of specific actions reinforced the personal rejection. For example, one confederate offered a cigarette (if *S* was a smoker) or chewing gum to the other, but pointedly neglected *S*. In the *Implicit Rejection*, there was silence during this period, unless *S* initiated conversation, in which case the responses of the confederates were minimal and neutral.

When *E* returned, he reluctantly announced that time prevented any further discussion of the topic; that it was a closed issue. He added that he would like to have them make a few ratings which would be helpful in interpreting the discussion. Following this rating procedure, *S*s were asked whether they would like to return for a discussion of the second topic with the same group. Assurance was given to each *S* throughout the postdiscussion procedure that his answers would be treated anonymously and confidentially. No *S* would know how the other responded. After collection of these data, *E* informed the *S* of the nature of the experiment, exposed the dissimulation, and requested that he refrain from discussing it with others.

RESULTS

Adequacy of Experimental Operations

Two questionnaire items may be examined for evidence that the manipulations had the desired effects. In one case, *S* is asked to make a quantitative rating of the degree he believes he is liked by the other *S*s. Responses were recorded on a 20-unit scale ranging from "dislike me very much" on the low end to "likes me very much" on the high. The second item requests a rating of how valid he thinks the other *S*s regard his opinions as expressed in the discussion. Responses on this 20-unit scale could range from "highly invalid" on the low end to "highly valid" on the high.

---

[2] There were actually four individuals who played the two confederate roles approximately the same number of times in each experimental condition.

Table 1 contains the mean ratings of the two items under the experimental conditions. It is clear that the experimental treatments had a significant effect on both measures. Rejected *S*s felt less well-liked personally and felt that their opinions were thought to be less valid than nonrejected *S*s. It can also be seen that while the means drop regularly across conditions, consistent with the assumption of a quantitative continuum of rejection, the differences between the Implicit and Explicit Rejection conditions are not statistically significant. Subsequent analysis will deal, consequently, with the Implicit and Explicit *S*s combined into a single Rejection condition.

Self-Depreciation

As hypothesized above, an irreversible rejection can be made more acceptable by taking the blame for it, that is, by lowering the self-evaluation in some degree in some circumscribed area. Two sources of data were assumed to be relevant to the hypothesis. One is a listing of positive and negative personality characteristics that all but a few *S*s provided following the questionnaire. The specific measure of this "personal" self-evaluation consists of the number of positive attributes minus the number of negative ones. The second measure of Self-depreciation is based on a quantitative rating by *S* of the validity of his opinions expressed in the discussion. Responses were recorded on a 20-unit scale.

TABLE 1

MEAN RATINGS ON MEASURES OF REJECTION OPERATIONS

	No Rejection ($N = 32$)		Implicit ($N = 32$)		Explicit ($N = 31$)	F	p (one-tail)
S's perception of how others[a] evaluate his opinions	20.4	$t = 2.68$ $p < .01$	16.6	*ns*	14.6	8.45	<.001
S's perception of how others[b] like him personally	14.7	$t = 7.84$ $p < .01$	7.1	*ns*	6.8	41.9	<.001

[a] Ratings of each confederate combined.
[b] Ratings of each confederate combined.

TABLE 2

MEAN RATINGS ON MEASURES OF SELF-DEPRECIATION

	No Rejection ($N = 21$)	Rejection ($N = 41$)	t	p (one-tail)
Positive minus negative personal self-evaluations	+3.1	+1.7	1.97	<.03
S's evaluation of his own opinions	($N = 32$) 16.8	($N = 63$) 14.9	3.13	<.005

Table 2 contains the average discrepancy between *S*'s positive and negative self-evaluations, and the mean ratings of the validity of his opinions under the two experimental conditions. It is evident that the average number of *S*'s positive self-evaluations as well as the estimate of the validity of his opinions were lower in the Rejection than in the No Rejection condition. Thus, whether the evaluation of himself is in personal terms or in terms of the worth of the opinions he holds, rejection definitely produces Self-depreciation.

Hostility

An external attribution of responsibility hypothesized to follow from rejection can be reflected in many forms of social aversion or hostility. Several measures were developed for different aspects of this general response tendency. To get at personal hostility, *S*s were asked to rate how much they liked or disliked each of the two confederates. A more impersonal measure of hostility called for *S* to rate the validity of each of the confederate's opinions. As an indirect measure of social hostility, *S* rated his amount of interest in the discussion issue. All of these ratings were on 20-unit scales. Finally, the most general measure of hostility was a quasi-behavioral one—a statement of whether or not *S* would commit himself to return for another discussion with the same discussion partners.

Table 3 shows the mean ratings on the three scales and the percentage of *S*s who express a willingness to return, under the two experimental conditions. All four measures assumed to represent hostility were significantly affected by rejection. Not only were the rejectors' opinions held to be less valid, but personally the rejectors were more disliked than the discussion partners were liked by *S*s in the No Rejection condition. Moreover, *S*s who had been rejected found the discussion topic less interesting than the *S*s in the No Rejection condition. Finally, for whatever reason, many more rejected *S*s were averse to returning for another discussion. In all, there is little doubt that the *S*s who were implicitly or explicitly rejected find their social environment relatively unattractive.

We turn now to a consideration of whether the responses produced by the experimental conditions are in any way related to each other within the same individual. Table 4 contains the product moment coefficients of correlation among the responses under each experimental condition. With respect first to the Rejection condition, there are two trends worth noting: (*a*) self-evaluation, in terms of the validity of one's opinions, tends to be positively correlated with each of the three measures of Hostility, and (*b*) liking others is positively correlated with one other measure of Hostility, namely, the evaluation of the opinions others hold.

TABLE 3

MEASURES OF VARIOUS FORMS OF HOSTILITY

	No Rejection (N = 32)	Rejection (N = 63)	t	p (one-tail)
Personal liking of others'[a]	21.3	18.3	2.21	<.03
S's evaluation of others' opinions[b]	28.4	17.6	4.76	<.001
S's interest is discussion	15.4	13.8	1.96	<.03
S's commitment to return	69%	50%	4.11[c]	<.03

[a] Ratings of the two confederates combined.

[b] Ratings of the two confederates combined.

[c] Chi square with 1 *df*.

Of greatest theoretical interest are the significant positive relationships between one of the Self-depreciation measures and two measures of Hostility. To determine whether the experimental rejection had anything to do with this result, the correlations under the No Rejection condition may be examined. It will be noted that the correlations between Validity of Own Opinion and each of the three Hostility responses, like their counterparts under the Rejection condition, are positive (two of the coefficients are at least at the .10 level). It thus seems fairly clear that a positive relationship between Self-depreciation and Hostility does exist for some measures, but is not due to the effects of experimental rejection.

In fact, when the differences between the correlation coefficients are tested for significance (by way of the *z* transformation and the *SE* of the difference between *z* scores), none of them is significant. In other words, no correlation seems to have been affected by the experimental rejection. Thus, the relationship between liking others and the evaluation of their opinions is significantly positive under the No Rejection condition as it was under the Rejection condition. It should also be noted that although the difference between experimental conditions falls short of significance, two of the correlations in which

TABLE 4

CORRELATIONS AMONG RESPONSES PRODUCED UNDER REJECTION AND NO REJECTION CONDITIONS

	Valid own Opinion		Like Others		Valid Others		Interest	
	No Rej.	Rej.	No Rej.	Rej.	No Rej.	Rej.	No Rej.	Rej.
Personal Self-Eval.[a]	+.39	−.09	−.02	−.05	−.03	−.15	+.39	−.05
Valid own opinion			+.36	+.46*	+.26	+.18	+.43*	+.28*
Like others					+.43*	+.37*	−.12	+.21
Valid other opinion							+.12	+.03

* These coefficients are at least of the size required for the .05 level of significance based on both tails of the *t* distribution.

[a] Note that correlations involving this measure are based on N of 21 and 41 in the No Rejection and Rejection conditions respectively. All other measures are based on Ns of 32 and 63.

Personal Self-evaluation is involved tend to be strongly positive ($p < .10$) under the No Rejection condition. There is no ready explanation of this, assuming that it is not merely a chance pattern.

DISCUSSION

There seems little doubt that rejection produces an increase in the strength of both self-depreciation and hostility in the sample of *S*s employed. Such findings are consistent with the hypothesis that an unchangeable rejection tends to bring about self-adaptive attributions of causality. Since, however, the experiment cannot prove the hypothesis unequivocally, it becomes necessary to summarize the arguments that support the hypothesis and oppose alternative interpretations.

Considering the finality of the rejection, the clear impossibility of communicating with the rejectors, and the anonymity of the responses it may be assumed that the behaviors produced by rejection did not primarily have a socially adaptive function. Thus, concretely, it appears to us to be inadequate to interpret the hostility merely as inhibited aggression toward the rejectors. Similarly, under the circumstances, self-depreciation could hardly have been produced in order to gain the rejectors' sympathy and thereby to rectify the damage to the self. On the other hand, it seems more plausible to suppose that the manifestly irreversible rejection strongly inclined the *S* to adjust himself to the negative situation rather than to try to change the situation. Although presumably there are other kinds of self-adaptive responses, hostility and self-depreciation appear to represent attributions of responsibility most closely. Clearly more definitive criteria of self-adaptive causal attributions are needed for further confirmation and extension of the theory.

It is important to point out that there were sources of devaluation anxiety in the experiment in addition to the rejection itself. For example, the discussion topic of premarital sex relations undoubtedly was a source of strong anxiety in many *S*s under both experimental conditions.[3] It is quite probable that the shame engendered in this way produced self-depreciation and even some hostility. In addition, there is the possibility that the very articulateness and urbanity of the confederates in agreeing and disagreeing with the *S*s made some of the latter feel extremely inadequate. This state of affairs might very well have created anger toward the confederates and attendant guilt, which in turn could have produced some self depreciation. Here again, however, granting our "rationalistic" classification of responses into socially and self-adaptive categories, there would not seem to be much of a basis for such "retaliatory" anger.

Given the motivation to make the self-adaptive responses of hostility and self-depreciation in both experimental conditions, what accounts for the fact that seemingly alternative self-adaptive responses tend in some cases to be positively correlated? One compelling possibility may be mentioned. In the concrete experimental situation of the *S*, it was not at all clear who was responsible for the rejection. It may be supposed that when the causal situation is basically uncertain, persons who are motivated to make self-adaptive responses tend to "spread" the responsibility over the major possible causes. Such a tendency would make for positive correlations among the responses.

SUMMARY

The experiment was designed to test the hypothesis that irreversible rejection gives rise to attributions of responsibility for that rejection, toward the self (self-depreciation) and toward the rejectors (hostility). The theoretical basis for the hypothesis was that personal devaluation anxiety engendered by the rejection can be made less unacceptable to the self through a causal analysis of it. The data were consistent with the hypothesis. A correlational analysis disclosed that a positive correlation existed between one of the self-depreciation responses and two hostility responses under both conditions of Rejection and No Rejection. The discussion pointed to sources of devaluation anxiety other than rejection, and to the supposed tendency for the person to spread responsibility in a situation

---

[3] Concerning this topic, the breakdown "For" and "Against" was approximately the same in all experimental conditions.

where the cause of the anxiety is relatively unclear.

REFERENCES

BANDURA, A., & WALTERS, R. Dependency conflicts in aggressive delinquents, *J. soc. Issues*. 1958, **3**, 52–65.

DITTES, J., & KELLEY, H. Effects of different conditions of acceptance upon conformity to group norms, *J. abnorm. soc. Psychol.*, 1956, **53**, 100–107.

PEPITONE, A. Attributions of causality, social attitudes and cognitive matching processes. In R. Taguiri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*, Stanford: Stanford Univer. Press, 1958.

WOLBERG, L. The character structure of the rejected child. *Nerv. Child*, 1944, **3**, 74–88.

(Received April 1, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

STIMULUS POOLING AND SOCIAL PERCEPTION[1]

RICHARD H. WILLIS[2]

University of Wisconsin

THE purpose of this experiment was to study the effect of stimulus pooling on judgments of personal attractiveness, where, in this context, stimulus pooling refers to the assembling of stimulus components into composite stimuli. An understanding of the relations between judgments made to composite stimuli and judgments made to components presented alone is of fundamental importance for the psychology of judgment. It has, in addition, important implications for social perception, e.g., for the development of social stereotypes and the formation of impressions of others.

Very briefly, the method employed was first to scale identification photographs on personal attractiveness, and then to assemble these photographs into sets of two or three. These sets, or composite stimuli, were then judged for overall attractiveness by additional judges.

One particularly simple kind of relationship that might be expected to hold between measures of judgments made to composite stimuli and those made to the stimulus components presented alone is that the former are equal either to the mean or the sum of the latter, depending upon the kind of task set for judges. Following Gulliksen (1956), this "linearity" or "additivity" of pooling effects may be derived by starting with the assumption that the rate of change of subjective value, v, with respect to an objective stimulus measure, x, is constant. This corresponds to the differential equation

$$dv/dx = k \qquad [1]$$

Upon integration and setting $x = x_0$ when $v = 0$, we obtain

$$v = k(x - x_0) \qquad [2]$$

For stimulus components i and j, and composite stimuli ij,

$$v_i = k(x_i - x_0) \qquad [3]$$

$$v_j = k(x_j - x_0) \qquad [4]$$

$$v_{ij} = k(x_i + x_j - x_0) \qquad [5]$$

By eliminating x_i and x_j between Equations 3-5, we have

$$v_{ij} = v_i + v_j + kx_0 \qquad [6]$$

Since scaling data give, not the subjective values, v, but scale values, s, which differ from the former by a constant, we substitute $s_i - C = v_i$, $s_j - C = v_j$, and $s_{ij} - C = v_{ij}$. The resulting expression is

$$s_{ij} = s_i + s_j + (kx_0 - C) \qquad [7]$$

Because judges in the present experiment were requested to respond to the *mean* of component scale values, rather than to their sum, we rewrite Gulliksen's equation (1956) as follows

$$s_{ij} = \left(\frac{s_i + s_j}{2}\right) + K \qquad [8]$$

This indicates that a plot of the scale values of composite stimuli against mean scale values of stimulus components will be linear with intercept $K = \frac{1}{2}(kx_0 - C)$ and—note—unit slope. When $kx_0 - C = 0$ in Equation 7, or $K = 0$ in Equation 8, we may speak of *additivity*, whereas the term *linearity* may be reserved for the more general relationship.

Gulliksen (1956) working with food preferences and using the method of pair comparisons, found Equation 7 to fit his data as well as three more complex functions, viz. logarithmic, square root, and negative exponential. Rimoldi (1956) working with "interest 'in knowing' the person or persons" and using three different methods of scaling, found correlations between obtained scale values of composite stimuli and those predicted from an additive relationship (Equation 8 with $K = 0$) to be .93, .97, and .96.

---

[1] This paper is based on a dissertation submitted to the graduate school of the University of Wisconsin in partial fulfillment of the requirements for the PhD degree. The writer wishes to express his sincere appreciation to J. C. Gilchrist, under whose supervision this investigation was conducted, and to David A. Grant and Leonard Berkowitz, who served as members of the author's doctoral research committee.

[2] Now at the Carnegie Institute of Technology.

On the other hand, what is known about the nature of social stereotyping suggests that under certain conditions, judgments made about composite stimuli (social groups) may not be so simply related to the mean judgments made about stimulus components (individual members of the groups involved).[3]

Campbell (1956), for example, has derived a principle of "enhancement of contrast" from the concept of stimulus generalization, whereby the prediction is made that the extent of overlap between two groups will be underestimated at intermediate levels of learning, due to a reduction of the variances of judgments within a group. Generalization among members of a group causes those relatively high on the continuum to be underestimated, while those relatively low are overestimated. This prediction was verified experimentally, using nonsense syllables. Clarke and Campbell (1955) found a slight degree of such relative exaggeration of group differences for white students' estimates of the examination scores of their Negro and white classmates.

However, Campbell's derivation applies to judgments of individual group members, not to judgments of the groups as such. A study which demonstrates the operation of enhancement of contrast in judgments of groups is that of Kephart (1954). He found a marked tendency for policemen to exaggerate the Negro-white crime rate differential in their districts. Interestingly, the higher the actual arrest rate in the district, the greater the tendency towards overestimation. This fact by itself does not imply that pooling effects are necessarily nonlinear, but it does imply that Equations 7 and 8 are not sufficiently general. In order to obtain the kind of enhancement of contrast found by Kephart, a slope greater than unity is required.

So far as the author is aware, no prior investigation has made a direct comparison of judgments of composite stimuli and those of stimulus components and at the same time has employed stimulus material somewhat analogous to that upon which judgments of social groups and their members is based. The main question to which the reported research was directed may be formulated as follows:

Are the processes involved in responding to composite stimuli by themselves sufficient to yield the exaggeration of group differences so often found in social sterotypes?

In the absence of any precise knowledge of the weighting laws involved, it seemed desirable to include as independent variables, in addition to differences in means of composite stimuli, such "intragroup" factors as heterogeneity and skewness of component scale values about composite stimulus means. Because interest was centered about pooling effects, it seemed obligatory to include as an independent variable the number of components per composite stimulus. Presumably any effects due to stimulus pooling, such as enhancement of contrast, should be accentuated by increasing the number of components pooled. Certain other variables, such as sex of judges, sex of photographic subjects, and order of presentation, were introduced primarily as experimental controls.

And finally, one additional variable, the number of judges working together, was included for the purpose of testing a specific hypothesis. It was expected that pairs of judges working together, making a single, collective response to each stimulus, would produce distributions of judgments with less dispersion than those produced by individual judges. It was hypothesized that a process of social accomodation would occur, whereby partners would reconcile differences in judgments by compromising, thereby avoiding extreme judgments.

---

[3] The situation is complicated—as, for example, Allport (1954) has pointed out—by the fact that social stereotypes often derive, not only from contact with group members, but also in large part from attitudes assimilated from nonmembers, especially those belonging to the individual's reference groups. Furthermore, systematic distortion of memory traces, so well documented by Bartlett (1932), plays an important role in the development of stereotypes. An analysis of such effects, however, is outside the scope of the present paper.

METHOD

Subjects. The *S*s were 212 University of Wisconsin undergraduates recruited from elementary psychology classes during the academic year 1956–57. Thirty-two *S*s served as judges during the scaling of the stimulus components, while 180 *S*s participated in the judging of the composite stimuli.

Scaling the stimulus components. Approximately 1,500 photographs, 1¼ in. square, of students entering the University in the fall of 1950, were obtained. From this initial group, two subgroups of 200 each were selected of males and females. Each subgroup included relatively high proportions of extremely attractive and extremely unattractive individuals.

Values on a scale of personal attractiveness were obtained for each of the 400 photographs by means of the method of equal-appearing intervals. The judges, 16 males and 16 females, used a sorting rack containing nine categories labelled A through I. A front cover on the rack prevented judges from observing the number of photographs placed in each category, thus increasing the independence of judgments. No restrictions were placed on the number of photographs to be placed in each category. Instructions (cf. Willis, 1957) were read by the judge, and then *E* explained the main points of the task. One of the two groups of 200 photographs was sorted first, then the other. Order of presentation was balanced both for male and female judges.

Experimental design: Overview. The entire experiment, subsequent to the scaling of components, may be considered and analyzed either as three $2^6 \times 4$ orthogonal factorials,[4] or as two interlocking $2^6 \times 4$ orthogonal factorials. In all, there were eight factors, of which seven had two levels and one (heterogeneity) had four.

Assembling the composite stimuli. Once the scale values of attractiveness were obtained, photographs were assembled into composite stimuli, or sets, and mounted on 3 × 5 index cards. Each set consisted of either two or three photographs (2-sets and 3-sets, respectively). Sets were systematically varied with respect to the following variables: (*a*) *set size*—the number of photographs per set, (*b*) set *mean scale value* (MSV)—the mean attractiveness scale value of set components, (*c*) *heterogeneity*—mean deviation of scale values about MSV, (*d*) absence or presence of *skewness* of the distribution of scale values about MSV, and (*e*) *sex of photographic subjects*, all photographs in a set being of the same sex.

Assembling the subseries and series. A number of sets were assembled into a *subseries*, a subseries being the sets given to a judge to sort at one time. All sets within a subseries were of the same size and sex. Each judge was required to sort matching male and female subseries, and order of presentation was balanced for each group of judges. There were three pairs of matching subseries, each pair consisting of a male and a female subseries. Each pair of matching subseries formed a *series*. Series I contained 2-sets. Series II contained symmetrical 3-sets, and Series III contained highly skewed 3-sets, where the measure of skewness was twice the intermediate scale value minus the sum of the other two. Within each MSV condition, skewness was always in the same direction; in the high MSV condition, the most extreme component in each set was also the one with the highest scale value, whereas the most extreme component in each low MSV set was the one with the lowest scale value.

Within each series, sixteen *set types* were represented. These set types were formed by the various combinations of male and female photographic subjects, high and low MSV, and four levels of set heterogeneity. Each set type in Series I was represented by six sets. In Series II and III, each set type was represented by four sets, with the exception of two set types which were represented by two sets. The six subseries were all matched, set type by set type, on both MSV and heterogeneity. In general, it was possible to achieve excellent matching.

Obtaining the judgments of composite stimuli. Sixty judges were used with each series, each judge being used with only one series. In half the sortings, a single judge responded to each set, using the same sorting rack used in the scaling. In the other half of the sortings, two judges of the same sex worked together in an informal way, making decisions about the placement of sets into the sorting rack. Judges working alone read a set of instructions which were identical with that used in the scaling, except for the addition of an instruction to take all photographs of each set into account in making judgments. Instructions read by partners were identical to those read by set judges working alone, except for an additional item instructing partners to alternate in the actual handling of sets, while deciding together about the placement of each set. *E* reviewed the important points in the instructions for judges before they began, just as was done during the scaling.

Sixteen scores were obtained from each *J*,[5] one for each set type. The score for a set type was the mean score for all sets of that type, while the score for a set was merely the number of the response category into which it was placed, categories being numbered 0 to 8, from least to most attractive.

---

[4] There were two error terms, Error Term I (MS, *J*s within conditions), and Error Term II (MS, residual). Error Term I was used to test the significance of differences between independent groups, while Error Term II was used with tests based on possibly correlated data, i.e., tests of sources of variation involving one or more of the factors of heterogeneity, MSV, and sex of photographic subjects—and thus also involving repeated measures. The *df* for Error Term I was equal to 32 times the number of series involved, while the *df* for Error Term II was equal to 480 times the number of series involved.

[5] The word *judge* will hereafter be used to designate the individual, either working alone or with a partner, while the symbol *J* will be reserved to refer to the judging unit, which may consist of either one person or two. A *J* corresponds to a subject for purposes of statistical analysis.

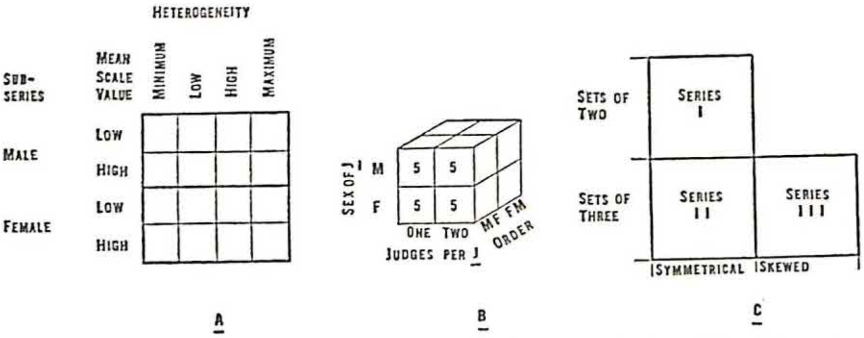


FIG. 1. The design of the experiment: the sixteen set types within a series, the eight subgroups of *J*s within a series, and the relationships between series with respect to set size and presence of skewness within sets.

Experimental design: Conditions. For each series, *J*s were balanced with respect to sex and number of judges per *J*. Order of presentation of subseries was also balanced. The entire design, which is diagrammed in Fig. 1, was replicated five times. In Fig. 1A are shown the sixteen set types that occurred in each series. There were eight experimental conditions for each series, each containing five *J*s, as shown in Fig. 1B, making 120 *J*s in all. Fig. 1C diagrams the interrelations between series. Series I and II are both free from skewness within sets, but differ in set size. Series II and III have set size in common, but the sets in Series III are skewed to the full extent allowed by the specified heterogeneity. Series I and III thus differ both in size of set and presence of skewness. All three series were matched on each of the six factors shown in Figs. 1A and 1B. Each series, taken separately, forms an orthogonal design in these six factors. Series I and II taken together form an orthogonal design in these six factors plus set size, while Series II and III taken together form an orthogonal design in these six factors plus skewness. Heterogeneity, MSV, and sex of photographic subject are within-*J* factors; number of judges per *J*, sex of judges, and order of presentation exist between conditions within series; and set size and skewness are between-series factors.

RESULTS

Overall main effects. In Table 1 are shown the subgroup means obtained when data from all three series[6] were subdivided by levels of each factor in turn. The results of the tests of significance are reported in Table 2 (See Footnote 4). Heterogeneity, set MSV, and sex of photographic subjects, tested against Error Term II, are all highly significant. None of the remaining overall main effects approach significance.

Set heterogeneity in moderate degree showed no effect, but mean set judgments were slightly lowered in the condition of maximum heterogeneity. A Duncan Range Test (Duncan, 1955) indicated that the three highest means were not significantly different from one another, but that the lowest mean was significantly lower than the next lowest mean, at the .01 level. Consequently, a "cluster-straggler" conclusion is accepted. Two facts stand out in Table 3, the second more interesting than the first:

1. Mean set judgments are slightly higher

TABLE 1

MEAN JUDGMENTS OF SETS WITHIN CONDITIONS[a]

Heterogeneity of set components	
Minimum	3.43
Low	3.34
High	3.39
Maximum	3.04
Set mean scale value	
Low	2.02
High	4.58
Sex of photographs	
Male	3.34
Female	3.26
Number of judges per *J*	
One	3.26
Two	3.34
Sex of judges	
Male	3.35
Female	3.26
Order of presentation	
M-F	3.26
F-M	3.35
Set size	
Two	3.26
Three	3.38
Skewness	
No skew	3.38
Skew	3.26

[a] Based on all data, with exceptions, specified in Footnote 6 in which comparisons were based on data from two series only.

---

[6] There are two exceptions: the set size and the skewness comparisons were based on data from two series only (Series I and II, and Series II and III, respectively), to avoid confounding the set size and skewness effects within a comparison. See Fig. 1C.

TABLE 2

Significance Tests of Main Effects[a]

Source of Variation	MS	df	F	p
Heterogeneity	5.88	3	10.45	<.001
Set mean scale value	1055.50	1	1876.40	<.001
Sex of photographs	4.59	1	8.16	.005
Number of judges per J	9.15	1	1.40	>.20
Sex of judges in J	7.83	1	1.20	>.20
Order of presentation	3.77	1	—	—
Set size	4.19	1	—	—
Skewness	4.13	1	—	—

[a] Based on all data, with exceptions, specified in Footnote 6 in which comparisons were based on data from two series only. Error Term II was used to test heterogeneity, set mean scale value, and sex of photographic subjects, while the other main effects were tested against Error Term I. See Footnote 4.

TABLE 3

Comparisons with Respect to Mean Scale Value and Mean Set Judgments

	Series I		Series II		Series III		Overall	
	(Level of MSV)		(Level of MSV)		(Level of MSV)		(Level of MSV)	
	Low	High	Low	High	Low	High	Low	High
MSV	2.14	4.15	2.14	4.15	2.16	4.16	2.15	4.15
MSJ	2.10	4.42	2.04	4.71	1.91	4.61	2.02	4.58

than the mean of the scale values of the set components.

2. The sets of low MSV were judged below their MSV, while the sets of high MSV were judged above their MSV. The drop of the low sets was 0.13 and the rise of the high sets was 0.43, so that the gap between the two increased from 2.00 to 2.56, or about 25%.

The overall difference between MSV and mean set judgment was significant beyond the .001 level, as indicated by a t test. Additional t tests indicated that the drop in the judgments of low sets, the rise in the judgments of high sets, and the increase in the gap between conditions were all significant beyond the .001 level. It is not likely that these effects could be due to a biased selection of set components from the full scale, for the grand mean for the scale values of all 400 photographs was 2.95, which is not far from the mean MSV of 3.15 for all sets. Furthermore, the direction of this difference would favor a downward shift, contrary to the results obtained. Slightly more than half of the total number of component scale values on the full scale were between 2.15 and 4.15, the levels for the low MSV and the high MSV conditions, respectively, so the increased gap between conditions cannot be accounted for by the possibility that too small an initial difference in MSV levels was selected.

Since judgments of sets are systematically different from the mean of judgments of set components, we conclude that the condition of additivity, as Johnson (1955, pp. 306–308) calls it, which has also been referred to as "the linear hypothesis" by Rimoldi (1956), does not hold for stimuli of the type used. Johnson and Rimoldi refer specifically to the situation described by Equation 8 with $K = 0$, but because of the pronounced enhancement of contrast found, we may conclude that Equation 8 does not hold, no matter the value of K, for enhancement of contrast implies a slope greater than unity.

The remaining significant overall main effect is that for sex of photographic subjects. Sets of males were rated slightly higher than sets of females. Further results, discussed below, indicate that the interaction of sex of photographic subjects with sex of J is more pronounced than the main effect.

Analysis of variance. The data were subjected to extensive treatment by analysis of variance (cf. Willis, 1957). Briefly, the analyses consisted of (*a*) an analysis of Series I and II taken together, (*b*) an analysis of Series II and III taken together, and (*c*) an analysis of each of the series separately. The purpose of the first of these analyses was to test the significance of the set size factor and its interactions with certain other variables. The second analysis was performed in order to make similar tests about the skewness factor and its interactions. A primary advantage of the series-by-series analyses is that they make it possible to take into account the consistency of results running through three independent groups of Js, an important consideration. In fact, it seems reasonable to give as much or more weight to consistency as to level of significance per se. Because of the large number of F ratios computed (63 per series), very conservative standards were adopted for accepting a finding as significant. The null hypothesis was rejected only for those sources of variation which were significant at or beyond the .05 level for *all three* series.[7]

The analyses of variance gave additional support for the significance of the main effects for heterogeneity and mean set value, but neither the main effect due to set size, nor that due to skewness, approached significance. Two interaction effects were found to be highly significant, however. First, the Series I–II analysis revealed that the difference in mean set judgments between the low and high MSV conditions is larger for 3-sets than for 2-sets, as shown in Table 4. This effect is significant beyond the .001 level. Not surprisingly, the effects of stimulus pooling were greater in the condition in which the larger number of stimulus components were pooled. It should be noted that the range of component scale values, both within high and low MSV conditions and overall, were very nearly the same for all three series, thus eliminating the possibility that differences in ranges could account for the obtained results.

TABLE 4

SUBGROUP MEANS FOR INTERACTION BETWEEN MEAN SCALE VALUE AND SET SIZE[a]

Set Size	Mean Scale Value	
	Low	High
2	2.10	4.42
3	2.04	4.71

[a] Based on Series I and II.

Second, the series-by-series analyses showed that both male and female *J*s tended to be more lenient towards their own sex (Table 5). This pattern of differences held for all three series, reaching the .01 level of significance in Series I and III, and the .05 level in Series II. The differential is a little more than twice as large for male *J*s as it is for female *J*s.

TABLE 5

SUBGROUP MEANS FOR SEX INTERACTION BETWEEN SEX OF PHOTOGRAPHIC SUBJECT AND SEX OF JUDGES[a]

Sex of *J*s	Sex of Photographs	
	Male	Female
Male	3.48	3.21
Female	3.20	3.32

[a] Based on all data.

Effects of number of judges per J on variance of judgments. In order to investigate the hypothesis that judgments made by partners exhibit less variance than those made by individual judges, mean variances were computed for the two *J* sizes. The mean variance for single judges was 37.97, while that for partners was 33.97. The difference, although in the expected direction, is not significant ($t = 1.20$, $df = 118$, $p > .20$). Pooling of responses did not produce a significant degree of convergence of judgments. Thus, even "the only consistent finding of significance for social psychology," cited by Johnson (1955, p. 471) in his discussion of group judgment, is not invariably found.

DISCUSSION

The most interesting results, from the point of view of social stereotyping processes, are (*a*) the highly significant enhancement of contrast, which resulted in an increased scale separation in the set-judging condition, and (*b*) the more pronounced nature of this effect with the sets having the larger number of components. These findings strongly suggest that, quite independently of learning and memory effects, there is a purely judgmental process that contributes to enhancement of contrast.

The enhancement of contrast observed here, however, is of a different kind from that found by Campbell (1956). The judges in his experiment made judgments about group members, not about groups as such. He found that the amount of overlap between two groups was considerably underestimated, a result which he had predicted from considerations of stimulus generalization. Neither his prediction nor his result involved an increased separation between group means, such as was found in the present experiment. In fact, his analysis calls for a *decrease* in the difference in group means due to intergroup stimulus generalization, a prediction which was very roughly confirmed by his data. Thus, although intergroup stimulus generalization may have played some role in the present experiment, its effects, if any, were more than neutralized by effects operating in the opposite direction on the differences in group means.

---

[7] For those sources of variation which were significant at or beyond the .05 level for *any two* of the three series, the null hypothesis was *tentatively* rejected. See the original thesis (Willis, 1957).

What underlying mechanism, then, may account for this absolute exaggeration of group differences? One possibility is that it is the result of efforts on the part of *J*s to reduce the cognitive strain resulting from ambiguity by sharpening intergroup differences. It can be assumed that *J*s had more difficulty in forming definite impressions when judging sets than when judging individual components, and that this difficulty, or cognitive strain, was greater for 3-sets than for 2-sets. The intergroup sharpening, along with relative intragroup leveling, served to reduce the impression of ambiguity and the concomitant cognitive strain. We may hypothesize that a need to reduce cognitive strain may also be one of the important motivations in the formation of social stereotypes, as well as in other instances of sharpening, leveling, and enhancement of contrast occurring in social perception.

A second possible mechanism, not mutually exclusive with the one above, is that *J*s were making judgments in terms of estimated probabilities, so that increasing distances in either direction from some central point on the response continuum corresponded to increasing degrees of unusualness. Then, in the set-judging condition, component probabilities were componded in some way to yield a lower probability for the set. This relatively low set probability would evoke a response assigning the set to a position at relatively great distance from the central point on the response scale. To take an analogy from everyday experience, it would be judged somewhat unusual to see a woman six feet tall, but it would be judged more unusual to see two or three such women together. Unfortunately, the data at hand do not allow a choice between the two interpretations.

Whatever the psychological mechanism or mechanisms may be, it is easy to show what is necessary mathematically to remove the restriction of unit slope from Equations 7 and 8, and so allow either for the possibility of sharpening (slope > unity) or leveling (slope < unity) of group differences. By assuming that a different rate-of-change constant applies for each group size, we replace Gulliksen's single differential equation (Equation 1) by a set of equations. For groups of Size 1 and 2, we write

$$dv_1/dx_1 = k_1 \qquad [9]$$

$$dv_2/dx_2 = k_2 \qquad [10]$$

where subscripts identify group size, and v and x are, as before, subjective value and objective stimulus measure, respectively. Since x_2 equals the sum of the two x_1, Equations 3–5 become

$$v_i = k_1(x_i - x_0) \qquad [11]$$

$$v_j = k_1(x_j - x_0) \qquad [12]$$

$$v_{ij} = k_2(x_i + x_j - x_0) \qquad [13]$$

By algebraic operations similar to those used to arrive at Equations 7 and 8, the following expressions are reached.

$$s_{ij} = K_1(s_i + s_j) + K_2 \qquad [14]$$

$$s_{ij} = K_1\left(\frac{s_i + s_j}{2}\right) + K_2' \qquad [15]$$

where $K_1 = k_2/k_1$, $K_2 = k_2x_0 - (2k_2/k_1 - 1)C$, and $K_2' = \frac{1}{2}K_2$. Equation 14 applies when judges are requested to respond in terms of the *sum* of the component scale values ("How much would you like to have this portable phonograph plus this pen and pencil set?"), while Equation 15 applies when judges are requested to respond to the *mean* of the component scale values, as in the present experiment. When $K_1 > 1$, group differences are exaggerated, while when $K_1 < 1$, group differences are underestimated. For the data of Series I $K_1 = 1.15$, and $K_2' = -0.37$. For three components per composite stimulus, Equation 15 may be generalized to read

$$s_{ijk} = K_1\left(\frac{s_i + s_j + s_k}{3}\right) + K_2' \quad [15]$$

where $K_1 = k_3/k_1, K_2 = 2k_3x_0 - (3k^3/k_1 - 1)C$, and $K_2' = K_2/3$. For the data of Series II and III taken together, $K_1 = 1.34$, and $K_2' = -0.90$.

Another finding which may have implications for social stereotyping is the slight, but significant, drop in mean scale values produced by extreme heterogeneity of component scale values within sets. Because the effect is absent with moderate degrees of heterogeneity, predictions made on the basis of set means alone will be nearly as accurate as predictions taking into account individual component scale values; but because the effect appears with extreme component heterogeneity, knowledge of the distribution variance may, in extreme cases, contribute to predictive power. To see how this might apply to social stereotyping, consider the example—entirely hypothetical—that it is known that Jews do not differ from other groups on mean introversion-extraversion, but are more widely dispersed about this mean, with higher proportions both of loud, aggressive individuals and of quiet, retiring ones. Stimulus pooling effects could conceivably contribute to a heavier weighting at one end of the continuum, and in this way facilitate the formation of a group reputation, even though the group in question does not differ in the mean from other groups.

The greater leniency of *J*s towards the own-sex subseries is likely due largely to the fact that judges were rating the individuals portrayed as potential dates or mates, and consequently adopted more critical standards. A second possibility is that, when judging members of one's own sex, a process of group identification produced a tendency to judge less critically. Identification might also have caused judges to assign themselves a position on the scale, not necessarily consciously. Then, the more leniently a given judge judges his own-sex subseries, the higher a rating he implicitly assigns himself.

In our culture, attractiveness is accepted by both sexes as being more important in the case of young women than it is for young men. This suggests that the date-rating set would be more pronounced for males judging females, but that group identification would be more salient for females judging females. Because the difference in mean set judgments is more than twice as large for male *J*s as it is for female *J*s, it seems probable that the first process is the more important one.

Turning to the negative finding of principal interest, the failure to find a significantly smaller variance of judgments in the "partners" condition, the most probable explanation is the lack of any but arbitrary, ambiguous scale units. This situation allowed partners, as compared to individual judges, to shift to a different equivalence ratio between the stimulus scale and the response scale. Partners no doubt felt as much of an obligation as did individual judges to use more than a few of the response categories, and because of this they expanded the scale just about the required amount to compensate for the effect of response-pooling. Presumably, if the stimuli had been of a sort that more clearly defined scale units (if scale units had been inches or dollars, for example), and if the relationship of responses to stimuli had been more clearly indicated (as with anchors), the reduction in variance would have been more pronounced. Experimentation to test this interpretation is currently in progress.

SUMMARY

The effects of stimulus pooling on judgments of personal attractiveness were investigated in an orthogonal factorial design, using as stimulus components identification type photographs which had been previously scaled on attractiveness. Composite stimuli were constructed from these components by mounting sets of two or three photographs on cards. These composite stimuli were varied systematically along several dimensions. Judgments of sets were then obtained under conditions similar to those under which components had been scaled. The most clear-cut and firmly demonstrated positive results were:

1. Mean set judgments (MSJs) were not equal to mean scale values (MSVs) of set components—the additivity condition did not hold. Rather, for sets of high MSV, $\overline{\text{MSJ}} > \overline{\text{MSV}}$, while for sets of low MSV, $\overline{\text{MSJ}} < \overline{\text{MSV}}$.

2. This enhancement of contrast was significantly more pronounced for sets of three components than for sets of two.

3. Extreme heterogeneity of component scale values slightly lowered MSJs; moderate heterogeneity had no significant effect.

4. Judges were more lenient in judging the attractiveness of own-sex photographs than other-sex photographs.

Negative results worthy of note were:

1. Judges working in pairs produced MSJs which were not significantly different, either in mean or variance, from those made by judges working alone.

2. The main effect due to set size was not significant.

3. The main effect due to the amount of skewness of component scale values about set MSV was not significant.

The implications of the findings for (*a*) social perception, especially social stereotyping, and (*b*) group judgment were considered.

REFERENCES

ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass: Addison-Wesley, 1954.

BARTLETT, F. C. *Remembering.* Cambridge, England: Cambridge Univer. Press, 1932.

CAMPBELL, D. T. Enhancement of contrast as composite habit. *J. abnorm. soc. Psychol.*, 1956, **53**, 350–355.

CLARKE, R. B., & CAMPBELL, D. T. A demonstration of bias in estimates of Negro ability. *J. abnorm. soc. Psychol.*, 1955, **51**, 585–588.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

GULLIKSEN, H. Measurement of subjective values. *Psychometrika*, 1956, **21**, 229–244.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

KEPHART, W. M. Negro visibility. *Amer. sociol. Rev.*, 1954, **19**, 462–467.

RIMOLDI, H. J. A. Prediction of scale values for combined stimuli. *Brit. J. statist. Psychol.*, 1956, **9**, 29–40.

WILLIS, R. H. Stimulus pooling and response pooling in social judgments. Unpublished doctoral thesis, Univer. of Wisconsin, 1957.

(Received April 13, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

ACHIEVEMENT MOTIVATION IN FANTASY AND BEHAVIOR[1]

DONALD M. BROVERMAN, EDWARD J. JORDAN, JR.,[2] AND LESLIE PHILLIPS
Worcester State Hospital and Institute of Human Development, Clark University

Evidence that motives can influence the content of perception has led to attempts to measure motivation by means of projective fantasy. McClelland, a leader in this movement, has presented a technique of assessing such motives as need Achievement, need Affiliation, and need Power from responses to modified versions of the TAT. Experimental work (McClelland, Atkinson, Clark, & Lowell, 1953; Atkinson, 1958) has indicated positive relationships between these fantasy measures and indices of the same motives derived from overt behavior. Such results have been taken as evidence for what might be termed a general-expression model of motivation, i.e., the view that motives are concurrently reflected in both fantasy and overt behavior.

Recently, however, the validity of this formulation has been questioned. Lazarus, Baker, Broverman, and Mayer (1957) found that achievement fantasy decreased with increasing age and education. Application of a general-expression model to this result implies a parallel reduction in the behavioral expression of achievement motivation with increasing age and education. This implication appeared implausible to Lazarus et al. Hence, these authors proposed that fantasy might serve as an alternative or substitute channel for the expression of achievement motivation when this motive is blocked from behavioral expression in real life. The higher achievement fantasy in younger, less well educated *S*s was attributed to their inability to express achievement needs in behavior, while the older individuals were seen as more able to express such needs, with a concomitant reduction of fantasy about achievement.

This alternative-channels model leads to the prediction of an inverse relationship between achievement fantasy and behavioral attempts to achieve. It thus stands opposed to the general-expression model of McClelland et al. (1953) which predicts a positive relationship between these two indices of achievement motivation.

The present study is concerned with an experimental resolution of these conflicting models. In attempting such a resolution, it seemed to us that the kind of behavioral index of motivation to be used was of critical consequence. Most previous experimentation has used for this purpose performances on laboratory experimental tasks, but in the light of a recent discussion of "task relevance" by Vogel, Baker, and Lazarus (1958) we felt that the assessment of achievement motivation in behavior by such means was of questionable validity. Laboratory tasks such as solving anagrams, solving arithmetic problems, or crossing out letters on a page would appear to have little relevance to an individual's principal interests in life, and might well yield data more influenced by such extraneous factors as the individual's proneness to believe laboratory instructions or his need to please the examiner. Furthermore, this technique subjects all *S*s to the same task, disregarding individual differences in specific areas of interest. Even if some *S*s were highly motivated by a laboratory task of arithmetic problems, for example, many others probably would not be. Finally, many experiments have intentionally used tasks that are relatively simple to perform, so that intelligence would not confound presumed effects of achievement motivation in performance. However, evidence (Lewin, Dembo, Festinger, & Sears, 1944) suggests that striving declines with relatively easy tasks, paralleling a decrease in level of aspiration.

---

[1] This investigation was supported by the Dementia Praecox Research Project, Worcester State Hospital and a research grant (M-896) from the National Institute of Mental Health, Public Health Service.

Thanks are due Melvin Feffer, Worcester State Hospital, and Morton Wiener, Clark University, for their many constructive criticisms. Stanley Kaden and Jerome Podell served as judges for the scoring of the Worcester Scale of Social Attainment as used in this study. Role-taking-task material was gathered and scored by V. Gourevitch. We should also like to thank Thaddeus Clapp through whose generous cooperation subjects were made available for use in this study.

D. M. Broverman participated in this research as a United States Public Health Service Postdoctoral Fellow in Clinical Psychology.

[2] Now at Veterans Administration Hospital, Durham, North Carolina.

The preceding considerations impelled us to measure the behavioral expression of achievement motivation by means of life history data, which appear much more likely to reflect behaviors toward goals that are important to the individual. Life history data are unique to each individual, obviating the possibility that an irrelevant task is being imposed on *S*. Furthermore, their use allows the individual to report on his attempts to achieve goals at a level of difficulty suitable to his achievement motivation.

The present study investigates the relationship between achievement fantasy and behavioral indices of achievement motivation taken from the area of employment, a central concern for our *S*s, who were adult males. A positive relationship between achievement motivation in fantasy and behavior would be taken as supporting the general expression viewpoint on this motive, while a negative relationship would support the alternative-channels formulation. In either event, the use of a behavioral index drawn from life rather than laboratory behavior was seen as lending greater weight to any evidence that might be obtained.

METHOD

Subjects

Thirty-seven members of a local church were used in this study. All *S*s were paid for their participation, but their pay went to the support of their parish. The sample was characterized by a wide range with respect to socioeconomic variables. Education varied from 5 to 19 years, with the median at 13 years. Annual income extended from $3,000 to $10,000, with a median of $5,000. Job level, as defined by the Dictionary of Occupational Titles (DOT), ranged from managerial and professional to semiskilled, with a roughly even distribution of *S*s in all job categories. Age ranged from 24 to 50 years, with a median of 36.

Procedures

*S*s were administered the Worcester Scale of Social Attainment (WSSA) in a group. In a later session individual inquiry into responses was conducted. The WSSA is a broad assessment of the degree to which an individual meets certain standards of accomplishment in society, and is discussed more fully by Phillips and Cowitz (1953). Two sections of this scale were selected for use in this study: Striving for Advancement and Job Level.

Striving for Advancement (WSSA, Item 7) taps the extent to which an individual actively seeks to better his employment position. Scoring on this item is based on answers to the following check list.

1. I would like to go to school to get a better job.
   (*a*) Do you have a particular school in mind?
   (*b*) Name of school?
2. I have visited plants looking for a better job.
3. I am interested in another job if it is offered.
4. I read the help-wanted ads every day.
5. I have asked my friends to look for jobs for me.
6. I am going, or have gone to school lately for job improvement.
7. Sometimes I think about looking for something better.
8. I have written letters to employers.
9. I am keeping my eyes open.
10. I have tried to get interviews.
11. I would like to have a better job.
12. I would like to learn a trade.
13. I expect to work for myself.
14. I would like to work for myself.

Responses to these questions permit classification of *S*s into following four categories:

1. Desire and intense activity, e.g., letters written about jobs; interviews applied for; contacts for better position actively sought.
2. Desire and tentative activity, e.g., is interested in another job if offered; does not seek a job; does not write although may look through ads; has vague long range goals.
3. Desire but no activity, e.g., says he would like a change but does nothing.
4. No desire and no activity.

*S*s were placed in one of these four categories by two independent judges, whose judgments correlated .98 (Pearson rank order correlation).

Behavioral manifestation of the achievement motive is typically understood as a kind of striving measure, or a measure of how much effort is involved in trying to achieve; it does not necessarily reflect actual achievement. At the same time, however, it seems possible that actual achievement may influence the amount of achievement fantasy produced by an individual. For this reason, *S*'s occupational level was used as a control for actual achievement in the area in which his behavioral expression of achievement motivation was examined. This control measure was taken from the Job Level item of the WSSA (Item 1). This item assigns scores to each level of occupation, as defined by the DOT. Higher scores are assigned to higher job levels.

The fantasy measure of achievement motivation was taken from the Role-Taking Task (RTT) devised by Feffer (1959). This is a projective test, based on Shneidman's Make-a-Picture-Story (MAPS). Three background settings were used: a factory scene, a living room scene, and one other chosen by *S* from the MAPS materials. A variety of figures (men, women, and children) were provided for *S* to place in these scenes. *S*s then told a story, using at least three figures, for each situation. After completion of the three stories, *S*s were presented with each situation again and instructed to retell their stories from the viewpoint of each character in the scene.

RTT stories are scored for motivational content by several criteria. Four of these were used in the present study, since they were McClelland's need Achievement categories of Achievement Imagery (AI), Stated Need for Achievement (nAch), Anticipatory Goal States (GA+, GA−), and Affective States (G+, G−). In a sample of 120 TAT stories presented and scored by McClelland et al., this portion of the need Achievement score correlated .95 with total need Achievement. Hence, it seems reasonable to assume that scores based on these four categories of need Achievement are highly representative of total need Achievement, as scored by McClelland et al. (1953). Each *S*'s achievement fantasy score is the sum of themes in his RTT stories which satisfy the criteria of any of the above four classes.

RESULTS

Achievement fantasy scores were analyzed in a two-by-two factorial analysis of variance. Independent variables used were: Striving for Advancement (WSSA, Item 7) and Job Level (WSSA, Item 1). *S*s in the two higher striving categories were compared with those in the two lower categories. Job Level was divided at the median into high and low groups. The requirement of proportionality between the four cells (Lindquist, 1953) necessitated the random elimination of three *S*s from the appropriate cells. The resultant *N* of this analysis is 34. Correlation between Striving and Job Level was negligible ($r = .11$).

Table 1 reveals that Striving is reliably associated with fantasy-expression of achievement motivation ($F = 4.35$, p less than .05). Table 2 indicates that the direction of this relationship is inverse: high strivers show low achievement fantasy, and low strivers show high achievement fantasy.

Table 1 further indicates that Job Level and fantasy-expression of need Achievement are not significantly related.[3]

TABLE 1
ANALYSIS OF VARIANCE OF ACHIEVEMENT FANTASY
($N = 34$)

Source	*df*	*MS*	*F*
Striving	1	79.53	4.36*
Job Level	1	4.63	
Interaction	1	4.99	
Within Error	30	18.22	
Total	33	SS = 635.88	

* *F* significant beyond the .05 level, two-tailed test.

TABLE 2
MEAN ACHIEVEMENT-FANTASY SCORES OF HIGH AND LOW STRIVING GROUPS

High Striving ($n = 17$)	4.11
Low Striving ($n = 17$)	6.47

DISCUSSION

This study demonstrates an inverse relationship between behavioral striving and achievement fantasy,[4] and thus supports an alternative-channels view of the expression of achievement motivation. Interpretation of these results must take account of the numerous studies which support the general-expression model by demonstrating positive relationships between achievement fantasy and striving. The alternative-channels view can no more account for these instances of positive relationships than the general-expression model can account for the present results. We are inclined to attempt a resolution of this conflict of results via an extension of the alternative-channels model.

We would approach this problem by suggesting that persons unable to express their achievement needs in life-situations will be impelled to seek alternative modes of need expression, not only in fantasy, as suggested by Lazarus et al. (1957), but also in behavioral striving when tasks simpler and less ambiguous than life problems are encountered.

The experimental tasks used to assess behavioral striving typically are safer and/or better structured than the situations in which most life goals are embedded. Also, because they are made easy, intelligence presumably plays little, if any, part in achievement on them. Therefore, persons who are blocked in expressing achievement motivation in real life might be expected to show a great deal of striving on such tasks. On the other hand, *S*s who are able to express their achievement needs in life should have no need to exert themselves on tasks so distant from real-life concerns. Such persons might be expected to show little striving in the laboratory. It then follows from the present finding, i.e., a negative relationship between life striving and achievement fantasy, that low laboratory strivers would be the low achievement fantasiers, while high laboratory strivers would be the high achievement fantasiers. Thus, a reversal of striving patterns from life to the laboratory may account for the positive relationships previously observed between striving and achievement fantasy. Such a reversal occurs, according to this argument, because performance of laboratory tasks, like fantasy, may serve as a substitute expression of blocked achievement motivation.

---

[3] Two other possible indices of achievement motivation, educational level and annual income, also failed to predict achievement fantasy. While it is impossible to draw general conclusions from these findings, it appears that actual achievement was not a predictor of achievement fantasy in the present sample.

[4] Vogel, Baker, and Lazarus (1958) also report an inverse relationship between achievement fantasy and a behavioral index of achievement motivation (performance on the McKinney Reporting Task). However, this relationship was found only with fantasy indices obtained after a failure-stress condition; no relationship was found with fantasy obtained under neutral conditions. Hence, the relevance of this finding to the theoretical issue discussed in this paper is unclear.

The extended alternative-channels model, then, seems preferable to the general expression model for several reasons. It can account for both positive and negative relationships between striving and achievement fantasy and it is supported by evidence drawn from life history data. Finally, this model is generally consonant with many current developments in psychological theory: the psychoanalytic emphasis upon thought and fantasy as consequences of delayed action (Rapaport, 1951); evidence presented by the Lewinian school for substitute activity and cosatiation (Lewin, 1935); and demonstrations of the vicarious relationship of fantasy to action by developmentally oriented studies (Krus, Werner, & Wapner, 1953; Hurwitz, 1954; Kruger, 1954; Misch, 1954).

Nevertheless, it is clear that any attempt at a definitive statement about the relationship between achievement fantasy and striving would be premature at present. The extension of the alternative-channels model advanced in the present paper has not yet been empirically tested. Furthermore, additional factors such as cognitive controls (Klein, 1954; Lazarus et al., 1957) and environmental conditions (Birney, 1959), may influence the kind of relationships that are obtained.

One such factor, particularly pertinent to this study, is variation in fantasy-assessment technique. The RTT, on which our results are based, differs from the usual achievement-oriented TAT by allowing *S* some choice in determining the test stimuli and by requiring the retelling of stories from several viewpoints. While it is difficult to envision how these differences might reverse the direction of relationship between striving and fantasied achievement, this possibility cannot be ruled out, and our findings must be qualified in this regard.

In any event, the present study casts doubt on the unqualified assumption that achievement motivation is generally expressed and supports the conceptualization of fantasy and behavior as alternative channels for the expression of this motive.

SUMMARY

This study investigates the relationship of achievement fantasy to behavioral indices of achievement motivation. Conflicting views of this relationship have been presented. McClelland et al. (1953) have assumed that fantasy is a direct index of a generally expressed motive state and have supported their position with findings of positive relationships between fantasy and behavioral indices of achievement motivation. Lazarus et al. (1957) on the other hand, have conceptualized fantasy as an alternative, substitute channel for the expression of achievement motivation when behavioral channels are blocked. This latter view implies an inverse relationship between fantasy and behavioral expression of motivation.

The present experiment is an attempt to determine which of these views is applicable when behavioral indices of achievement motivation are derived from life history data.

Thirty-seven adult males, representing a wide range of age and socioeconomic variables, were administered the Worcester Scale of Social Attainment (WSSA) and a Role-Taking Task (RTT). Two behavioral indices of achievement motivation were taken from the WSSA: Job Level, a measure of actual achievement, and Striving for Advancement, a measure of the degree to which an individual is actively trying to raise his level of occupational achievement. A fantasy measure of achievement motivation was derived from the RTT.

An inverse relationship was found between the striving and fantasy measures of achievement motivation. No significant relationship was found between actual achievement and either achievement fantasy or striving.

These results were interpreted as supporting the alternative-channels model of Lazarus et al. (1957). Previous research in this area was criticized, and an attempt was made to extend the alternative-channels model to cover the reported instances of positive relationships between achievement fantasy and behavioral expression of motivation.

REFERENCES

ATKINSON, J. W. *Motives in fantasy, action, and society.* Princeton, N. J.: Van Nostrand, 1958.

BIRNEY, R. C. The reliability of the achievement motive. *J. abnorm. soc. Psychol.*, 1959, **58**, 266.

FEFFER, M. The cognitive implications of role-taking behavior. *J. Pers.*, 1959, **27**, 152–168.

HURWITZ, I. A. A developmental study of the relationships between motor activity and perceptual processes as measured by the Rorschach test. Unpublished doctoral dissertation, Clark Univer. 1954.

KLEIN, G. S. The personal world through perception. In R. R. Blake, & G. V. Ramsey (Eds.), *Perception, an approach to personality.* New York: Ronald, 1951.

KRUGER, A. Direct and substitute modes of tension-reduction in terms of developmental level: An experimental analysis by means of the Rorschach test. Unpublished doctoral dissertation, Clark Univer., 1954.

KRUS, D., WERNER, H., & WAPNER, S. Studies in vicariousness: Motor activity and perceived movement. *Amer. J. Psychol.*, 1953, **66**, 603–608.

LAZARUS, R. S., BAKER, R. W., BROVERMAN, D. M., & MAYER, J. Personality and psychological stress. *J. Pers.*, 1957, **25**, 559–577.

LEWIN, K. Substitute activity and substitute value. In *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald Press, 1944. Chap. 10.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MISCH, R. The relationship of motoric inhibition to developmental level and ideational functioning: An analysis by means of the Rorschach test. Unpublished doctoral dissertation, Clark Univer. 1954.

PHILLIPS, L., & COWITZ, B. Social attainment and reactions to stress *J. Pers.*, 1953, **22**, 270–283.

RAPAPORT, D. Toward a theory of thinking. In D. Rapaport (Ed.), *The organization and pathology of thought.* New York: Columbia Univer. Press, 1951.

VOGEL, W., BAKER, R. W., & LAZARUS, R. S. The role of motivation in psychological stress. *J. abnorm. soc. Psychol.*, 1958, **56**, 105–112.

(Received April 30, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

ATTITUDINAL CONSEQUENCES OF COMMITMENT TO UNPLEASANT BEHAVIOR[1]

JACK W. BREHM
Duke University

THE theory of cognitive dissonance (Festinger, 1957) has led to empirical investigation of the psychological consequences of inconsistencies between attitudes and behavior. It has been shown, for example, that a person induced to argue against his private opinion will tend to show change in his private opinion toward the induced position (Cohen, Brehm, & Fleming, 1958; Festinger & Carlsmith, 1959). It has also been shown that simply engaging in a behavior inconsistent with one's attitudes results in attitude change toward consistency with the behavior (Aronson & Mills, 1959; Mills, 1958). Other derivations concerning the consequences of inconsistency between attitude and behavior have yet to be tested. This paper reports a test of a relatively "nonobvious" derivation.

According to the theory, when a person is induced to engage in some behavior that he ordinarily would avoid, he experiences dissonance. He then tries to reduce the dissonance, and, in the present case, one way he may do so is by changing his attitude toward the behavior (or some aspect of it) in such a way as to reduce the reasons for avoiding it. Thus, an individual induced to perform a *disliked* behavior should tend, among other things, to *increase his liking* for that behavior.

The person's tendency to change his attitude is conceived to be a direct function of the magnitude of dissonance experienced. Given his dislike of a behavior and with the inducing force held constant, the magnitude of dissonance increases as the amount of committed behavior increases. And given the fact that he has committed himself to some disliked behavior, the less support that further information to which he is exposed gives to that behavior, the greater the magnitude of dissonance.

Specifically, then, the following derivations may be made.

1. The greater the amount of disliked behavior to which a person commits himself, the greater the magnitude of dissonance and consequent increase in liking for the behavior or the behavioral object. This hypothesis has been tested by Aronson and Mills (1959), who showed that the more severe a person's initiation in joining a group, the greater his liking for the group members and their activities.

2. Given commitment to disliked behavior, the less supportive is additional information, the greater the magnitude of dissonance and consequent increase in liking. This effect is expected to increase as a function of the amount of committed behavior. The present study was designed to test this derivation.

The theory also asserts that a person experiencing dissonance tries to reduce it and to avoid increasing it. Thus, the extent to which information either for or against engaging in the behavior is believed should be a function of the amount of committed behavior. This consideration gives rise to another derivation tested by this study.

3. Supporting information is accepted and nonsupporting information is rejected by a person in proportion to the amount of disliked behavior to which he is committed.

METHOD

The general procedure has been described in a previous report (Brehm, 1959). Briefly, it consisted of (*a*) a pre-experimental measure to ascertain a vegetable disliked by each of the subjects (*S*s), (*b*) an individual experimental session in which each *S* was induced to commit himself to eating some of the disliked vegetable, (*c*) exposure of the *S*, after he was committed, to information in support of or against eating the disliked vegetable, and (*d*) measurement of relevant affect and opinions before and after the experimental inductions.

The *pre-experimental questionnaire* was given to eighth-grade students in their English classes.[2] It consisted of a response scale for each of 34 vegetables. *S* was to indicate how much he liked each vegetable by an X on any one of 61 dots running along a straight

---

[1] This project was conducted as part of the Yale Communications Research Program. It was supported by a grant to Carl I. Hovland from Bell Telephone Laboratories.

[2] For their help in conducting this study, the author wishes to thank Louisa Willard, Psychologist for the New Haven Schools, Frank Carr, Principal of Troupe Junior High School, Teacher Connolly, and the eighth grade teachers, especially Teacher Himmelfarb.

line and identified at intervals of 10 dots by the following labels: like very much; like a lot; like a little; neither; dislike a little; dislike a lot; dislike very much. *S*s' responses were then scored from +3.0 for "like very much" to −3.0 for "dislike very much."

Experimental procedure. Each *S* was assigned, during school hours, to go to the Homemaking room of the school where he was met by the experimenter (*E*). It was explained that the project was "consumer research" about various people's preferences for different kinds of vegetables. *E* said that he wanted to ask the *S* to answer some questions about a vegetable picked at random (though it really had been selected as one quite disliked, at least −2.0 on the pre-experimental questionnaire). After answering some preliminary questions, *S* was given the main questionnaire to fill out. The questions, to be described in detail, were designed to measure liking for the vegetable and opinions about its food value.

On completion of the main questionnaire, *S* was told that his answers would be very helpful but that a person sometimes gave different answers if he first tasted the vegetable. It was further explained that *S* need not do anything further, since he might not like the idea of eating some of the vegetable, but that he would be given a reward if he chose to eat the vegetable and answer further questions about it. At this point the experimental manipulation of amount of committed eating was introduced.

Commitment to little eating (Low Eating condition) was accomplished by telling *S* that to obtain the reward he must eat a single small dish of the vegetable immediately. Commitment to greater eating (High Eating condition) was accomplished by telling *S* that to obtain the reward he must eat a small dish of the vegetable immediately and also return three or four times within a few weeks, each time to eat more. All experimental *S*s were offered the same reward, a gift certificate good for their own choice of two 45 rpm or 78 rpm records, or two tickets for movies of their own choice. *S*s were assigned at random to the two commitment conditions. It should be noted that the overall rate of refusal to eat was only 5% and did not differ between conditions.

If *S* agreed to continue, he was given a small dish of the vegetable, heated if appropriate, and allowed a few minutes in which to eat it. When finished, the *S* was given a "research report that he might be interested in reading." It was ostensibly from the laboratories of the testing organization and purported to give the food value of the vegetable. Actually, each report was designed to give information in support of (Support condition) or against (No Support condition) eating the vegetable. The following report, for example, was designed to *support* the eating of spinach.

> Vegetables have two important values in a person's diet—they furnish most of the minerals and most of the vitamins needed for strong bodies and active minds.
>
> Our tests of spinach show that this vegetable is particularly strong on vitamins. An average size portion will furnish 95% of the daily quantity of vitamins needed for building new body tissues and mending injured tissues and bones. Spinach is probably the best single vegetable to supply vitamins.

Other reports in the Support condition were identical except for the name of the vegetable. Reports in the No Support condition were identical except that they said "weak" instead of "strong," "only 5%" instead of "95%," and "worst" instead of "best."[3] The type of report given to each *S* was determined by chance.

When *S* had finished reading the report he was asked to fill out again the main experimental questionnaire (identical to the "before" questionnaire). In making the request, *E* mentioned that some people change one way, some another, and some don't change at all. He also reaffirmed the anonymity of the *S*'s answers.

Upon completion of the questionnaire, *S* was allowed to choose his reward and was asked not to talk about what he had done when he returned to class. Shortly after the completion of the project, *E* met with all the *S*s in a group and explained the deceptions and true purpose of the project.

Control procedure. In order to estimate any direct effects of the communications without a commitment to eat, a classroom experiment was performed. For this test 8th graders from a different school had to be used.[4] The second school is not far from the first and draws its students from a population with similar socioeconomic status. The same preliminary questionnaire was given, though this time by the *E* and only one week prior to the experimental session. A set of materials was prepared for each *S*, and consisted of the main questionnaire under which was a sealed booklet made up of either a Support or No Support communication for a vegetable disliked by the *S*, and a second copy of the main questionnaire. These were passed out to the students in class and *E* explained he was trying to obtain more detailed information by asking each person several questions about one particular vegetable whose name he would find on his questionnaire. *S*s were asked first to tear off their names, which were on separate top sheets. The questionnaires, thus, looked anonymous although they were coded on the back. *E* then asked *S*s to fill out the first questionnaire carefully but not to unseal the booklet until told to do so. When the first questionnaires were completed and passed in, *E* explained that while this information would be good, he had also found that people sometimes gave different answers if they thought carefully about the vegetable first, trying to imagine as vividly as possible just how it tastes, smells, and feels in the mouth. So, he said, he was also asking them to do this and then to answer some further questions they would find on the inside of the booklet. Also, on the inside of the booklet they would find a laboratory report about their vegetable in which he thought they would be interested, and they should read that before answering the questionnaire. After all final questionnaires had been completed and turned in, the true purpose of the experiment was explained.

---

[3] About half of the *S*s received similar communications concerning mineral instead of vitamin content. Since this type of communication failed to affect liking either directly or in conjunction with commitment, the results for these subjects are omitted from this report.

[4] The author wishes to thank Louisa Willard, Principal Corrigan, Principal of Truman School, and Teachers Walker, Smernoff, and Kelly, at Truman School, for their help in conducting this part of the study.

The main experimental questionnaire. This questionnaire included one question identical to that used on the prequestionnaire, which read, "Considering everything about it, (taste, smell, appearance, texture) to what extent do you like or dislike this vegetable?" It was answered on a scale like that used on the prequestionnaire. A second question pertained to the food value of the vegetable, and read, "How high would you guess the general vitamin content of this vegetable compared to others?" The answering scale was similar to that used for the previous questions except that the identifying labels were: very high; fairly high; slightly above average; just average; slightly below average; fairly low; very low.

Results

Commitment and change in belief. While the content of the communications was direct and forceful and the source was highly authoritative, *S*s might still manage to believe them differentially. According to dissonance theory, the greater the magnitude of dissonance being experienced, the greater the seeking of information that can reduce dissonance and the avoidance of information that might, increase it. Thus, persons in the High Eating condition, more than those in the Low Eating or Control conditions, should tend to raise, or resist lowering, their estimate of vitamin content.

Mean changes in estimated vitamin content are given in Table 1. The means for all three conditions show a decrease in the No Support and an increase in the Support condition, the difference being statistically reliable for each. It is clear that the manipulation of Support and No Support was successful. Mean changes in the High Eating condition, however, show less reduction in estimated vitamin content from the No Support communication, significantly different from the Low Eating condition at the .07 level, and from the Control condition at beyond the .01 level.[5] Similarly, the increase in estimate from the Support communication is greater in the High Eating condition than in either Low Eating or Control conditions, though not significantly so. If the effects of the Support and No Support communications are averaged, the High Eating condition is significantly different from the Low at the .04 level, and from the Control at beyond the .01 level. It is apparent, then, that *S*s tended to believe supporting information and disbelieve nonsupporting information in proportion to the amount of eating to which they were committed.

TABLE 1

Mean Initial Scores and Changes in Estimated Vitamin Content

Commitment	Communication			
	No Support		Support	
	Initial	Change (N)	Initial	Change (N)
Control	.58[a]	−3.02 (11)	.84	0.92 (10)
Low Eating	.45	−2.35 (8)	1.76	0.77 (7)
High Eating	1.23	−0.34 (7)	.46	1.66 (10)

[a] Comparisons of the initial means do not yield significant *t*s, indicating that the initial scores are somewhat unstable with *N*s this small.

Effect of support on liking. The communications were given in a class room situation to estimate their effect on liking for the vegetable. It was assumed that any such effects obtained in the Control condition would be at least as large in magnitude for the individually tested *S*s is the main experimental conditions.

Mean changes in liking (from the "before" to the "after" main experimental questionnaire) were computed for the Control conditions. These changes, shown in the first row of Table 2, indicate that the Support communication produced increased liking, the No Support decreased liking, the difference between the conditions being significant at the .05 level.

Effect of commitment and support on liking. Since the Control condition showed communication *support* to increase liking while the dissonance hypothesis holds that *nonsupport* increases liking, the two effects are working against each other. Thus, in the Low Eating condition, where the magnitude of dissonance is low, we might expect the support manipulation to have a direct effect on liking, similar to that in the Control condition. But in the High Eating condition, where the magnitude of dissonance is greater, the effect of attempts to reduce dissonance should reduce or even reverse the direct effect of the support manipulation on liking.

Mean liking change scores for each experimental condition are presented, along with the Control data, in Table 2. The direct effect of

---

[5] All p values are from two-tailed t tests.

TABLE 2

MEAN INITIAL LIKING AND CHANGES FOR EACH SITUATION

Commitment	Communication			
	No Support		Support	
	Initial	Change	Initial	Change
Control	−.52[a]	−.43[b]	−1.56	.46
Low Eating	−.99	.11	−.66	1.97
High Eating	−.03	1.33	−.64	.48

[a] None of the below comparisons yields a significant *t* for initial scores.

[b] *p* values for various comparisons of change scores:
Control (No Support-Support) = .05
Low Eating (No Support-Support) = .06
High Eating (No Support-Support) = .09
No Support (High Eating-Control) = .02
No Support (High Eating-Low Eating) = *NS*
Support (Low Eating-Control) = <.01
Support (Low Eating-High Eating) = <.01
Low Eating (No Support-Support)–High Eating (No Support-Support) = .02
Control (No Support-Support)–High Eating (No Support-Support) = .03

the support manipulation occurs in the Low Eating condition: increase in liking is greater for the Support than for the No Support communication (significant at the .06 level). And as expected from the dissonance hypothesis, the difference between Support and No Support conditions with the High Eating manipulation is reversed. While the difference is significant at only the .09 level, it is significantly different from that for the Low Eating condition at the .02 level, and from the Control condition at the .03 level. It is clear, therefore, that the direct effect of supporting information on increased liking tends to be reversed to the extent that there is behavioral commitment.

It will also be noted in Table 2 that commitment has no over-all effect on liking. The Low Eating conditions produced as much increased liking on the average as did the High Eating. These results therefore fail to provide adequate support for the hypothesis that increased liking is a simple function of amount of eating committed.

DISCUSSION

This experiment has shown that a person induced to perform a disliked behavior tends to increase his liking for the behavior as a joint function of the amount of the behavior to which he is committed and the presence of further supporting or nonsupporting information. The major results are in support of Festinger's theory of cognitive dissonance. When a person is induced to perform a disliked behavior, additional information can decrease or increase the magnitude of dissonance by supplying information that is respectively for or against engaging in the behavior. He attempts to reduce the dissonance, in proportion to its magnitude, by increasing his liking for the behavior. Thus, he increases his liking for a behavior to which he is committed to the extent that he is exposed to information against engaging in the behavior.

While the predicted interaction was obtained, the manipulation of amount of committed eating failed to produce an over-all effect on liking. Average change in liking in the Low Eating condition was as great as in the High. One possible reason for this unexpected result is that *S*s in the High Eating condition, who were expecting to return again, were more restrained in their general reaction to the situation. Another possibility is that the Eating manipulation was not sufficiently strong to provide an overall effect; only in conjunction with the No Support communication did it produce dissonance. The latter interpretation, however, would not explain why with the Support communication the Low Eating condition produced greater increased liking than did the High Eating condition. A third possible interpretation is that the pressure to change liking is a function of the discrepancy between behavior and information. That is, where behavior is minimal and informational support high, or behavior is great and support low, there is pressure to increase liking. When, however, both behavior and support are great, the pressure to increase liking is low. Thus, it is suggested that dissonance may be created by exposure to supporting information, and may be reduced interchangeably by commitment to behavior or by increased liking. This implication of the theory is not developed in Festinger's statement (1957).

The experiment, it should be noted, was concerned with only one possible channel of dissonance reduction. It assumed that attitude change toward the induced behavior would be a major mode of reduction under the present conditions, though other modes were available to subjects and may have been used to some degree. While this assumption was supported by the positive findings of the study, it might not prove to be a good assumption in other settings. As yet, dissonance theory does not completely specify the factors controlling the choice of channel through which dissonance is reduced. Making the theory operational therefore requires taking such extratheoretical factors into account. There is here a clear need for theoretical refinement and further research. Meanwhile, the results of the present experiment can be generalized to other situations only with due caution.

SUMMARY

This study tested a "nonobvious" hypothesis derived from Festinger's theory of cognitive dissonance: given the fact that a person is committed to an unpleasant behavior, he tends to increase his liking for that behavior more if he is exposed to information against engaging in it than if he is exposed to information favorable to engaging in it. This tendency is anticipated to be in proportion to the amount of such behavior to which he is committed.

The hypothesis was tested by inducing junior high school students to commit themselves to eating different amounts of disliked vegetables. Fictitious research reports to which they were exposed said their vegetable was either high or low in food value. Estimates of food value and liking for the vegetable were measured before and after the experimental inductions.

It was found that the communications about food value tended to be believed or disbelieved as a function of the amount of eating to which the person was committed. In addition, the main experimental hypothesis, as stated above, was supported.

REFERENCES

ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *J. abnorm. soc. Psychol.*, 1959, **59**, 177–181.

BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379–382.

COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, **56**, 276–278.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Row, Peterson, 1957.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

MILLS, J. Changes in moral attitudes following temptation. *J. Pers.*, 1958, **26**, 517–531.

(Received July 6, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

THE FREQUENCY OF SELF-REFERENCE STATEMENTS AS A FUNCTION OF GENERALIZED REINFORCEMENT[1]

J. STACY ADAMS AND BROWNING HOFFMAN
Stanford University

RECENT experiments have demonstrated applications of operant conditioning in the realm of verbal behavior (e.g., Hildum & Brown, 1956; Verplanck, 1955; see Krasner, 1958, for review). Stimuli demonstrated to be reinforcing have been verbalizations of assent, agreement and approval, nods, smiles, paraphrases and impersonal events such as buzzer and bell tones and colored lights. Attention, although hypothesized by Skinner to be a generalized reinforcer (Skinner, 1953, 1957) has not been specifically investigated.

The purposes of this study were (*a*) to determine whether a combination of attention and assent constituted a reinforcing event, (*b*) to explore the effects of prolonged extinction on the occurrence of self-reference statements and other behavior such as emotional reactions and escape behavior, and (*c*) to determine what are the implications of the use of common events (attention-assent) in free, unstructured interview situations. In most previous studies extinction has not been carried out at length. In Verplanck's investigation, for example, extinction was carried out for only 10 minutes, a period too brief to uncover any effects other than those upon the class of operants considered (Verplanck, 1957).

METHOD

General Experimental Plan

The experiment was conducted under the guise of being a survey of student opinion with two groups of 31 *S*s, one experimental group and one control group. The survey interview was divided into four consecutive periods, the first three each lasting 10 min., while the last varied in length up to a maximum of 20 min. Each period was further subdivided into 2-min. intervals during each of which *S*'s self-reference statements and other statements were counted. After the introduction of the "survey" and after *S* understood what was wanted of him, the first 10-min. period began and the operant level of statements containing self-references was determined, no reinforcement being given in either group.

During the second 10-min. period *E* reinforced all statements containing self-references made by experimental *S*s by following their occurrence by a combination of attention and assent. A precise description of reinforcing events is given below. Control *S*s received no reinforcement in Period II, nor in any other period.

Extinction was conducted during the third 10-min. period and during the subsequent 20-min. period. As will be seen, some *S*s stopped responding entirely during the course of extinction. Extinction consisted of withholding from experimental *S*s the reinforcement previously given in Period II.

If at any time during the course of the experiment *S*'s over-all rate of responding fell to zero, nondirective probes such as "What else can you think of?", "Can you think of anything else?", or "Do you have any other ideas?" were used. These were infrequently necessary except during extinction and then they were often unsuccessful in maintaining continued responding.

Subjects

The experimental *S*s were 31 male and female, graduate and undergraduate students at Stanford University, who were "interviewed" in the spring of 1958. Control *S*s were drawn from the same population but were "interviewed" in the spring of 1959. It was ascertained at the conclusion of the experiment, when the real purpose of the "interview" was explained, that all *S*s were naive to the extent that they had no training in conditioning and had never before participated in an experiment on operant conditioning.

Experimenters

The *E*s were advanced undergraduate and graduate students in two of the senior author's courses on language and verbal behavior. Seven of these collected the experimental group data; two collected the control group data. One *E* collected data from both groups. In addition to having some theoretical sophistication in verbal operant conditioning, each *E* had a minimum of four complete practice sessions followed by a detailed critique period. The *E* who collected data from both groups had, in addition, one year's intensive practice in verbal conditioning in connection with a number of other studies. He acquired considerable sensitivity in identifying statements and self-reference statements and imparted this skill to the other *E* who collected control data and for whose training he was responsible. As a consequence, to which reference will be made below, rates of response show an elevation in the control group data as compared to the experimental group data.

In view of the fact that there is a difference in the sensitivity of the experimental and control group *E*s, it would have been desirable to obtain interexperimenter reliability data. Unfortunately, any useful reliability data would have had to be based on the reinterview of the same *S*s by different *E*s and this, of course, would have had an effect on the observed responses. Our feeling is that *E*s of equivalent training would have yielded acceptably high reliabilities. Since the seven *E*s who collected the experimental group data had comparable training and the two who collected control group data had equivalent training, the inter-*E* reliabilities within the two groups are probably acceptable. And this is really the important point, for we are especially concerned with change or lack of change *within* each group.

---

[1] The assistance of Beverly A. Brown, George E. Dunn, Ben V. Hole, Frank B. Krasne, Lillian Lauer, Jack Lyle, and Aline Spivock is gratefully acknowledged.

Procedure

The experiment was conducted at the convenience of *S*s in their living quarters or other quiet place at a time when only *E* and *S* were present and when *S* had one full hour available. When the experiment proper began, *E* was seated in front of *S* and recorded the frequency of self-reference statements and all statements at the end of each 2-min. period.

In his introduction to *S*, *E* told him that he was doing a "survey" for the psychology department, that the purpose was to find out students' "frank and honest personal opinions and ideas about conformity and education in America," and that *S* could talk for about an hour or more on the subject. *S* was further told that *E* would not enter into the interview except to take notes. *S* was then reminded that the interview would be most fruitful if a complete, detailed and exhaustive expression of his ideas and opinions were obtained.

The topic "conformity and education in America" was selected for several reasons. First, it was a topic students frequently discussed and about which they held quite definite opinions. Thus, it was felt that there would be some motivation to express views on it. Secondly, the subject was broad enough to permit sustained talking. Thirdly, it was a topic which neither imposed an upper nor a lower limit on the use of statements of self-reference.

Except during the 10-min. reinforcement period for the experimental *S*s, *E* kept his entire attention on his notebook and was silent. He did not lift his head, shift his line of regard to *S*, or make any verbal responses, unless *S* stopped talking. When the latter occurred *E* used a general, nondirective probe in order to try to maintain *S*'s output.

Reinforcing Stimulus Events

A single class of stimulus events was hypothesized to have reinforcing properties. The event consisted of *E*'s (*a*) lifting his head, (*b*) shifting his line of regard from his notebook to *S*, and (*c*) murmuring "mm-hmm" simultaneously. After a reinforcement (which took about 1 sec.) *E* returned to his original position and composure until *S* emitted another statement containing a self-reference. As previously indicated, it is only among experimental *S*s in Period II that reinforcement of self-references was used. Among control *S*s, *E* was silent throughout the experiment and kept his attention on his notebook.

The Class of Operant Responses Conditioned

The verbal operants selected for conditioning were all statements containing one or more references to *S* personally. Thus, statements having "I," "me," "my," "mine," "myself" and like words were considered self-reference statements, except in those instances when the pronoun had someone else than *S* as a referent (e.g., "my roommate thinks . . . "). Statements containing the first person plural pronoun, "we," were also counted as self-reference statements if they pertained to an activity in which *S* himself had participated, or a belief he shared.

Statements were defined as units of "thought," containing a complete idea. Statements, therefore, were complete sentences as defined in standard dictionaries, although in speech they might not be clearly punctuated as a written sentence. Such a definition is not entirely satisfactory, for the sufficient and necessary criteria are not enumerated, and there is, therefore, room for subjective variability on the part of *E*s. *E*s, however, become more proficient with experience and there is a tendency for them to identify more self-reference statements as their proficiency increases.

RESULTS

Effects of Reinforcement

The basic index of responding used in the analysis is the ratio of self-reference statements to total statements made (called SR ratio hereafter). The mean SR ratios for each period and the mean SR ratio over-all periods are plotted in Fig. 1 for the experimental and control *S*s.

The first point to be noted is that the mean over-all SR ratio in the experimental group is .305, whereas it is .408 in the control group. The difference is significant beyond the .001 level and represents increased proficiency by one *E* in identifying statements and self-references as he gained experience during a period of approximately one year, as previously related.

As Fig. 1 shows, the size of the SR ratio increases during reinforcement and drops during extinction in the experimental group, as predicted under the hypothesis that attention-assent constitutes a reinforcing state of affairs. To determine the significance of the observed change in SR ratios two analyses were performed. First, Cochran's *Q* test (Siegel, 1956) was applied to determine whether the values of SR deviated significantly from the mean over-all SR value of .305. This required a determination of whether each individual ratio deviated above or below that value. The *Q* value

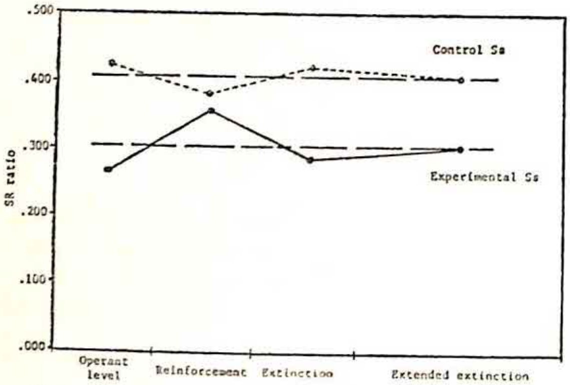


FIG. 1. SR ratio by period for experimental and control subjects.

obtained was 9.38, which is significant beyond the .05 level. Among the controls the SR values during each period did not differ significantly from the over all mean SR value of .408 ($Q = 2.68$).

The second analysis, Friedman's nonparametric analysis of variance (Siegel, 1956), reveals that the treatments contributed significantly to the variance among experimental *S*s ($\chi_r^2 = 16.7$) but not among control *S*s ($\chi_r^2 = 5.15$). Further analysis of the data for the experimentals revealed that the reinforcement treatment is the major contributor. Wilcoxon's signed-ranks analysis for matched pairs (Siegel, 1956) indicated that the SR ratio during reinforcement among experimental *S*s was significantly greater (*a*) than the SR ratio during the operant level period ($z = 2.5$), (*b*) than the SR ratio during the first extinction period ($z = 2.9$), and (*c*) than the SR ratio during the second extinction period ($z = 3.1$).

The rate of emitting self-reference statements and all statements is shown in Fig. 2. Among the experimental *S*s the rate of self-reference statements increases significantly from 1.42 to 1.90 self-reference statements per minute during the reinforcement period ($z = 2.98$, by Wilcoxon's test), while the rate of emission of all statements remains virtually constant. Among the control *S*s both the rate of self-references and the rate of all statements drops significantly during the corresponding period ($z = 3.20$ and $z = 3.13$, for self-reference and all statements, respectively).

Effects of Extinction

The SR ratio among experimental *S*s decreases significantly following reinforcement, during the first extinction period ($z = 2.9$). In the second extinction period, lasting up to 20 min., there is no further decrease in the SR ratio (Fig. 1). However, the rate of self-reference statements decreases continuously during extinction, as may be seen in Fig. 2. By Wilcoxon's test, the rate of self-reference statements during the first extinction period is significantly smaller than during reinforcement ($z = 3.4$) and is significantly smaller during the second extinction period than during the first ($z = 3.08$). The rate of making all statements is not significantly smaller during the first extinction period than during the reinforcement period ($z = 1.27$). However, the rate is significantly smaller during the second extinction period than during the reinforcement period, ($z = 4.3$). Furthermore, the rate during the second extinction period is smaller than during the first extinction period ($z = 3.96$). Therefore, although the SR ratio during the second extinction period is not smaller than during the first extinction, the rate of emitting self-references and total statements decreases during successive extinction periods.

The rates of self-reference and all statements for control *S*s do not vary significantly after the first 10-min. period of the "interview." The extinction effects noted among experimental *S*s are in no way evident among the controls. Hence, the observed experimental extinction effects are attributable to withdrawal of generalized reinforcement.

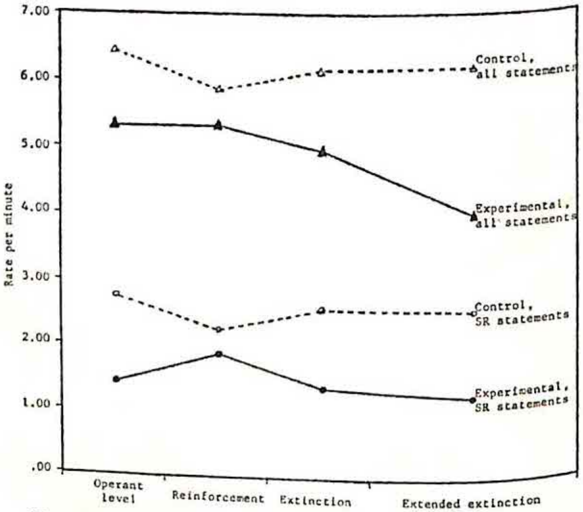


FIG. 2. Rate of emitting self-reference and all statements per minute for experimental and control subjects.

Many experimental *S*s displayed general disturbance, emotionality, and, in some instances, downright anger and hostility toward *E* during extinction. This was especially noticable in the late stages of extinction. In her report on her "interviews," one of the experimenters wrote: ". . . many *S*s found the interview to be a stress situation (during extinction)." This finding is in accord with reports by Verplanck (1955). There was further evidence that some *S*s "left the field"—i.e., refused to continue speaking, despite prompts—because the discontinuance of reinforcement was aversive to them. It is as though nonreinforcement constituted *punishment* for them. However, it is not possible to say that all *S*s whose verbal behavior stopped before the end of the second extinction period did so because of aversion. In only one *S* was there evidence of emotionality among the controls, though a number of them stopped responding before the end of the 50-min. experimental session, despite the use of probes, apparently having run out of comments on the subject of conformity.

In Fig. 3 the number of *S*s who continued to respond during each 2-min. interval of the extinction period is plotted. At the beginning of the extinction period all experimental *S*s were responding; during the corresponding period twenty-eight of the thirty-one controls were responding. The difference is not significant by Fisher's exact test ($P = .12$, one-tailed). At the end of the extinction period, however, nine experimental *S*s and sixteen control *S*s were responding. The difference is significant beyond the .05 level ($\chi^2 = 3.28$, one-tailed).

Evidence of Awareness

Upon questioning at the end of the experiment not one experimental or control *S* reported that the "interview" situation was other than was related to him. Some experimental *S*s reported "awareness" of change in *E*'s behavior during the reinforcement and extinction periods, this change being one of "interest" and "attention." None, however, were able to report on what these changes were contingent—i.e., emitting self-reference statements. We may conclude, therefore, that the observed experimental effects were not a function of knowledge or awareness of the experimental manipulations.

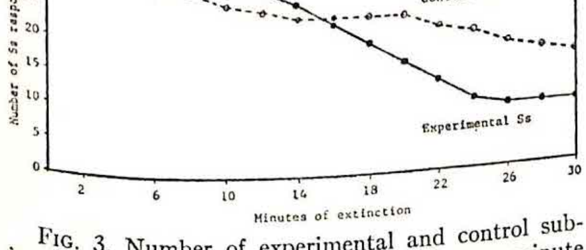


FIG. 3. Number of experimental and control subjects continuing to respond at various two-minute intervals of extinction.

DISCUSSION

The evidence is unequivocal that the frequency of emitting self-references can be controlled by making attention-assent contingent upon their occurrence. When attention and assent follow, self-references increase (relative to total output); when attention and assent are withdrawn, the relative frequency of self-references decreases. The same is true for the rate of emitting self-reference statements. These effects are not observable among the controls, whose self-reference statements were not reinforced at any time. Further evidence of the effects of the experimental manipulations comes from the fact that *all* verbal behavior is extinguished in significantly more experimental than control *S*s.

Four explanations of the latter over-all extinction phenomenon suggest themselves. First, it is plausible that *S*s stop responding because they have exhausted all that they can say on the subject of conformity and education in America. Though this is undoubtedly true in the case of some *S*s, since a number of control *S*s stopped responding, such an explanation does not account for the fact that the experimental *S*s are more likely than the control *S*s to stop responding before the termination of the experiment. Nor does it account for the experimental *S*s' greater emotionality during extinction. Second, it is possible that statements without self-references were extinguished along with those containing self-references because of generalization. Both types of statements are quite similar in that both are verbal responses and it is therefore possible that extinction of self-reference statements generalizes to other statements. However, in the present instance the explanation appears improbable because, if there were stimulus generalization, it should have been evident during reinforcement. But there was no evidence for stimulus generalization during reinforcement, for only the frequency of self-reference statements and rate of emitting them increased, while total output and rate of total output remained constant.

A third possibility is that a minimum number of self-reference statements is necessary if talking on the selected topic is to be possible. There is no direct evidence in support of this explanation. However, the data in Fig. 1 indirectly suggest that it may have validity. After prolonged extinction the SR ratio among experimental *S*s drops only to the operant level in those *S*s who continue talking. Thus, approximately .270–.300 may be the *minimum* possible ratio if talking is to continue. If the proportion of self-references drops below this level (because of extinction) talking may become impossible or at least very difficult. Further support for this hypothesis comes from the fact that while the SR ratio remains constant during extinction, the *rate* of self-references and total output steadily decreases. Among the controls, on the other hand, both the SR ratio and the rate of self-reference statements and of total output remain constant during the corresponding periods.

A fourth possible explanation of the extinction of all verbal behavior is that withdrawal of reinforcement constitutes punishment and that this punishment has the effect of arresting all verbal behavior (and perhaps other behavior) in the presence of *E*. If punishment is defined as "an operation in which an aversive or conditioned aversive stimulus is made contingent upon a response" (Ferster & Skinner, 1957, p. 731), the explanation seems reasonable, for silence is probably used by the verbal community as a conditioned aversive stimulus. As Skinner points out, remaining silent when the occasion usually gives rise to speech constitutes punishment (Skinner, 1957). Thus, in the present experiment the *E*'s withholding of attention and assent, following reinforcement, may have had an effect very similar to the silence following the commission of a verbal *faux pas* in a social group. That some experimental *S*s who stopped responding may have perceived the situation in precisely this way is evidenced by their reporting that they felt the *E* had become uninterested in or even annoyed at what they were saying and by the fact that they frequently became visibly tense, angry, and hostile.

Of the four possible explanations suggested, the data rule out that relating to the generalization of extinction. Running out of comments to make may partially account for the *S*s stopping to respond altogether, for some *S*s in the control group, as well as in the experimental group, stopped before 50 min. elapsed. However, it could not account for the significantly greater proportions of experimental *S*s who stopped. This disparity may be accounted for by either the third and fourth explanations offered, or, perhaps, by both. The data do not permit a choice between the two, however.

With respect to interviewing of the free, unstructured type, the findings of this study have clear implications. Attention and assent—both common interviewer responses—can produce significant changes in the content of a respondent's verbalizations. In this study, the effect was to increase the frequency of self-reference statements. When one member of a dyad differentially reinforces specific verbal response classes of the other participant, he gains some control over the conversation. Such control has obvious advantages. Psychotherapists may use verbal reinforcement techniques to guide their patients toward significant insights; these procedures may be of considerable importance, for instance, when applied to self-reference, affect, or other statements where therapeutic circumstances preclude direct requests for such information from the patient. Similarly, survey interviewers can control the content of a respondent's verbalizations in a direction consonant with the purpose of an interview by use of differential reinforcement. Or, a physician might find it useful to reinforce differentially in his patients verbal descriptions of symptoms, as opposed to diagnostic inferences.

Perhaps more importantly, interviewers should be aware of the possible undesirable influence which verbal reinforcement may have upon the responses of interviewees. Inappropriate use of such techniques may easily lead to selective biases in the data or clinical evaluations based upon reflected idiosyncracies of the interviewer which have been reinforced in a respondent or patient. Furthermore, the withdrawal of reinforcement, as in the present study, may result in emotionality that might have undesirable consequences, particularly in psychotherapeutic situations.

SUMMARY

An experiment was conducted to determine whether the use of a combination of attention and assent by an "interviewer" constituted a reinforcing event and to find out the effects of prolonged extinction. Following a 10-min. period during which the operant level of self-reference statements was measured, all self-reference statements by experimental *S*s were reinforced by *E*'s showing attention-assent for a period of 10 min. The *S*s' responses were then extinguished for a period of as long as 30 min. The same procedure was followed with control *S*s with the exception that they received no reinforcement.

It was found that attention-assent constituted a reinforcing state of affairs since the relative frequency and rate of self-reference statements significantly increased during conditioning among the experimental *S*s, while control *S*s showed no change. Extinction of self-reference statements produced a significant decrease in their relative frequency and rate of occurrence among experimental *S*s, but not among control *S*s.

During extinction most experimental *S*s showed emotional disturbance; in some cases anger and hostility. The verbal behavior of significantly more experimental than control *S*s was totally extinguished before the end of the 30-min. extinction period. In some instances it was evident that *S*s were leaving an aversive field. Possible explanations of the observed extinction phenomena were offered.

There was no evidence that any *S* was aware or had knowledge of the experimental manipulations.

REFERENCES

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

HILDUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, **53**, 108–111.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

SKINNER, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

(Received July 1, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

INTERPERSONAL PERCEPTION AND COMMUNICATION IN DISCUSSION GROUPS OF VARIED ATTITUDINAL COMPOSITION[1]

IRWIN ALTMAN[2] AND ELLIOTT McGINNIES
University of Maryland

IN A recent paper, the investigators reported a series of measures that distinguished between groups with homogeneous attitudes discussing a persuasive film on the topic of ethnic minorities (McGinnies & Altman, 1959). The present study applied these and additional measures to groups differing in attitudinal composition so as to determine the relationships existing between group structure and discussion behavior. Certain of Festinger's assumptions concerning forces to communicate in groups were also testable in the design of the experiment.

The structural properties of groups, and particularly the relative homogeneity of member orientations toward an issue, have been shown to influence perceptual and judgmental processes of the members (Sherif, 1932; Asch, 1952). In many of these studies, no systematic analysis of behavior was undertaken during the actual process of social interaction. We have attempted to follow influence pressures in different group structures throughout the process of member interactions, as well as to describe some relationships between group composition and interpersonal perception.

Recent studies (Festinger, 1950; Festinger & Thibaut, 1951; Festinger, Back, Schacter, Kelley, & Thibaut, 1950) have focused on the process of group functioning and have analyzed changes in member behavior as they influence and are influenced by others in the group. In general, these studies support the hypothesis of a positive relationship between degree of opinion discrepancy concerning an issue and the relative strength of pressures to communicate with those holding deviant opinions. In the majority of these studies, however, the relative balance of divergent opinions within groups has not been independently controlled. For example, some studies use four-man majorities facing one-man minorities, while others have much larger majorities. Since group composition has been shown to be an important determinant of member behavior, we controlled the degree of opinion discrepancy in groups by systematically varying majority-minority conditions.

Several additional differences between the present study and others dealing with intragroup pressures should be noted. Most studies in discussion settings have used situations where members have had to rely on others in the group for confirmation of their views. In the present study, certain members of the group found their views supported by the content of a persuasive film presented as a topic for discussion. In addition, earlier studies have defined opinion discrepancy as the degree of separation of members along an attitude continuum. Rather than manipulate the degree of distance between individuals in each group, we varied the number of individuals who represented extreme ends of an ethnocentrism continuum.

Briefly, the aims of the present study were as follows:

1. To explore the effects of various degrees of minority-majority conditions on member behavior during the course of group discussions.

2. To determine the relationships among attitudes, interpersonal perceptions, and forces to communicate within the different types of groups.

METHOD

Independent Variables

A slightly modified version of the California E scale was administered to 500 male students of the junior and senior classes at two high schools.[3] On the basis of their scores, *S*s were randomly assigned to one of five types of discussion groups, each of which consisted of six members. A total of 25 groups were formed, five in each of the following structures:

---

[1] This research was supported in part by a special grant from the National Institute of Mental Health, United States Public Health Service, Department of Health, Education, and Welfare, and is based upon portions of a doctoral dissertation submitted by the first author to the University of Maryland.

[2] Now with Human Sciences Research, Inc., Arlington, Virginia.

[3] We are grateful to the faculty and staff at Bladensburg and High Point Senior High Schools, Maryland, for their assistance in the conduct of this research. John Lesser, Robert Lana, and Richard Page assisted in the collection of data and management of the discussion sessions.

Low E homogeneous: All six members had low E scale scores.

Low E majority: Four *S*s were low Es; two were high Es.

Balanced groups: There were equal numbers of high and low Es.

High E majority: Four members were high Es; two were low Es.

High E homogeneous: All six group members were high Es.

Each group met on a single occasion 3–4 weeks after the initial attitude screening. The first part of the meeting was devoted to viewing a film dealing with the problem of minority groups. This film, "The High Wall," takes an antiprejudice position in dealing with the conflict between teenagers of different ethnic backgrounds. Following the film, the group discussed the general problem of ethnic conflict. The discussion was led by a member of the research team who assumed a nondirective role and stimulated rather than guided the discussion. During the discussion, the investigators monitored recording equipment and coded the sequence and direction of participant responses. From both these "spottings" and recordings, we were able to compute a number of indices of discussion behavior.

At the conclusion of each discussion, group members completed a questionnaire dealing with attraction to the group, attraction to particular group members, importance of the discussion topic, and perceived consensus of opinion. In addition, *S*s were readministered the E scale.

Dependent Variables

In order to assess the effect of various group structures on member behavior, three classes of dependent measures were used:

1. *Discussion Behavior*. Several discussion behavior indices previously developed (Gerard, 1953; Dickens, 1955; McGinnies & Altman, 1959) were employed. These measures required minimal inference by observers and were readily derived from the recorded discussions.

(*a*) *Spontaneity*. In discussion situations characterized by permissive leadership, opportunity is afforded group members to display varying degrees of initiative in their discussion behavior. The extent to which members act independently of leader stimulation is referred to as spontaneity. Operationally, a remark which occurred without leader probing was rated as spontaneous, whereas member responses which followed leader questions, probes, etc. were defined as nonspontaneous. A percentage index of spontaneity was obtained by computing the number of spontaneous remarks relative to the total number of responses.

(*b*) *Rate of response*. This general activity indicator is computed by determining the number of responses emitted by group members (or groups as a whole) per unit of time.

(*c*) *Spread of participation*. In discussions where there is no requirement that members participate, it is probable that inequalities in response output will occur. To compare the homogeneity of verbal output in the various structure conditions, a measure of spread of participation was employed (Dickens, 1955).

(*d*) *Direction of response*. Since we recorded the originator and the recipient of every discussion comment, it was possible to chart the direction of communication flow in the various groups. A modification of a measure developed by Gerard (1953) was used to represent response-direction scores.

2. *Attitude Change*. The E scale was administered several weeks prior to the film-discussion session and again immediately following the discussion. Although our major aim was to assess variations in discussion behavior and interpersonal perceptions as a function of group composition, we used this opportunity to relate discussion behavior to shifts in attitude.

3. *Questionnaire Data*. In addition to the E scale, a questionnaire consisting of measures of attraction to the group, sociometric preference, relative importance of the discussion topic, and accuracy of interpersonal perception, was administered at the conclusion of each discussion. Some of the more interesting and significant findings emerged from analysis of these data.

RESULTS

Discussion Behavior

(*a*) *Spontaneity*. This index, which refers to the degree of initiative of group members, was computed by dividing the number of spontaneous participations by the total number of participations during each of three equal time periods. Indices were computed for each group as a whole as well as for high and low Es separately. Since the meetings lasted approximately one-half hour, each index was based on about ten minutes of discussion.

A repeated-measures analysis of variance indicated significant changes in spontaneity over the course of the meetings for all groups, as well as for high and low Es considered separately. All groups exhibited more initiative as the meetings progressed (Fig. 1). Although no reliable differences in spontaneity appeared

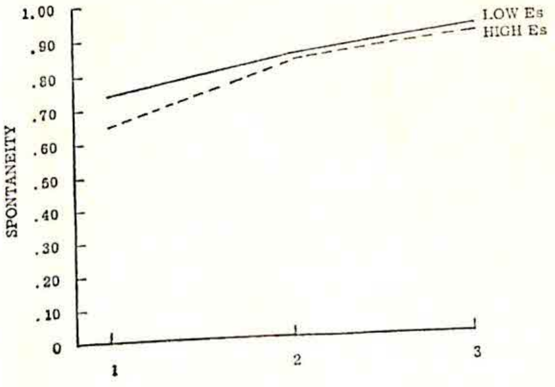


FIG. 1. Spontaneity indices over time periods.

among the various structure conditions, the low Es were generally more independent of the discussion leader than the high Es. These results are consistent with those obtained earlier by McGinnies and Altman (1959). In addition, there was a tendency for the balanced groups to exhibit lower spontaneity scores than any of the other groups, suggesting that structures balanced with respect to opinion heterogeneity lead to a reduced level of member initiative.

(*b*) *Rate of response.* A measure of discussion pace, or rate of output, was computed by dividing the number of participations in each of three time periods by the length of the time period and adjusting for the number of members in groups and subgroups. The index was computed for each total group, as well as separately for high and low Es.

Analysis of response rates showed no significant changes over the course of a meeting for either high or low Es. Moreover, no significant variations in discussion activity occurred for the different group structures. While statistical evidence is lacking, the data of Figure 2 suggest that all group members quickened the place of their discussion behavior as the meetings progressed. As shown in the figure, the low Es had generally higher rates of participation than the high Es ($t = 4.08$, $p < .01$). Thus, low Es tended to be both more active and less dependent on the leader than the high Es. It is interesting to note, however, that the rate of response difference stems largely from the heterogeneous structures. When there is absence of opinion divergence on the discussion topic, there is practically no difference in participation rates of high and low Es.

It should also be noted that the balanced groups again exhibited a unique pattern of discussion behavior. Rates of response for these groups were lowest in spite of the fact that neither attitude subgroup was dominant. Festinger's hypothesis (1950) might have predicted a maximum degree of both spontaneity and discussion pace in such groups.

(*c*) *Spread of participation.* To measure the relative equality of participations in the discussion, an index developed by Dickens (1955) was used. This measure, based on the relative discrepancy between each individual's response output and the group mean, yields an indication of the extent to which certain individuals dominated the discussions.

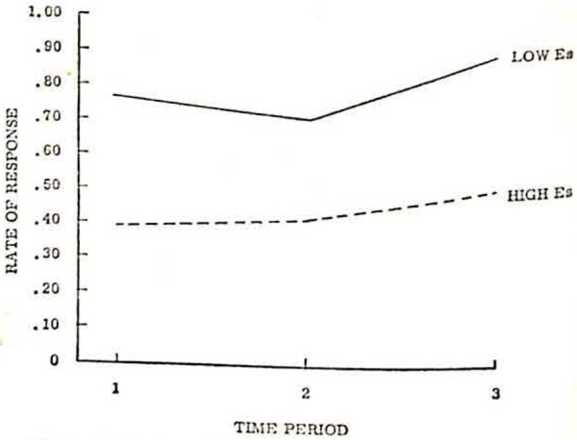


FIG. 2. Rate of response over time periods.

While Kruskal-Wallis analyses of variance showed no differences in relative spread of participation among the various group structures for either high or low Es, there were significant differences between the two attitude groups. Low Es tended toward equality in their individual contributions to discussions, while the high Es participation was dominated by a few individuals.

(*d*) *Direction of response.* The directional flow of communications between members was computed by dividing the number of responses directed toward a given subgroup by the product of the number of possible recipients, the number of members in the communicator subgroup, and the length of a time period. These D scores were computed for each of the possible communication channels in the heterogeneous structures: high E to low E; low E to high E; low E to low E; and high E to high E.

These scores were analyzed by Friedman's analysis of variance to ascertain temporal changes in communication flow and by the Kruskal-Wallis procedure to evaluate group structure and attitudinal differences in response direction.

When analyzed temporally, no differences in direction of communication flow occurred. High and low Es did not direct their remarks differentially over time. However, there were significant differences in communication flow for each of the four channels between the various structure conditions. The results show, first, that there was a paucity of opposition directed communication in the balanced groups. When no clearcut majority existed, members tended to direct their remarks largely to those participants who held views similar to their own. When considered with the rate of response and spontaneity data, we arrive at the picture of balanced groups being relatively lethargic or apathetic in their discussions.

A second set of results indicate that low Es tended to increase communication toward both high and low Es as their number decreased, i.e. as they moved toward the minority position. Conversely, high Es tended to decrease opposition-directed remarks as they progressed toward the minority position. Thus, there is evidence for the presence of an interaction between attitude and group structure as determinants of influence processes.

Attitude Change

Since the E scale was administered to *S*s several weeks before discussions and again after each meeting, it was possible to compare changes in attitudes across the different group structures. Analysis of covariance of post-discussion scores indicated that neither high nor low Es changed differentially in ethnocentric attitudes as a function of group structure. However, there was a significant decrease in ethnocentrism for pooled *S*s ($p < .01$).

Questionnaire Responses

In addition to the E scale, participants completed a short questionnaire following the discussions. The questionnaire covered a number of general topics including (*a*) the extent to which members were attracted to their group, (*b*) the personal relevance of the discussion topic, (*c*) the attraction of members to one another, and (*d*) the perception of opinions held by other members.

(*a*) *Attraction to the group.* Although low Es were least attracted to the balanced groups, they did not exhibit differential attraction to the various group structures ($p > .05$). On the other hand, high Es did show significant attraction differences, with the lowest degree of positive feeling occurring in the balanced groups. Thus, both types of *S*s were least positive in attraction to their groups in those structures where neither attitude subgroup was in the majority. One other general finding, though not significant, was that low Es seemed to be more frequently attracted to their groups than the high Es.

(*b*) *Relevance of the discussion topic.* Although there was a slight tendency for low Es to place greater importance on the discussion topic, neither predisposition group responded differentially to this item across the various group structures.

(*c*) *Sociometric choice.* By having each *S* rank the other members with respect to personal attraction, we obtained another measure of the extent to which *S*s perceived one another's attitudes as these were revealed in the course of group discussion.

Although analysis of variance showed no difference between the choice patterns of high and low Es, there were significant differences in choices among the various structure conditions and a significant attitude × structure interaction. Examination of the data shows that the interaction results from the fact that the low Es were more apt to choose other low Es in both the minority and majority positions, while the high Es showed no appreciable differences across structures. In other words, the low Es reacted sociometrically to differences in attitudinal composition of the groups. The high Es apparently were imperceptive of such differences, at least so far as sociometric relationships were concerned.

(*d*) *Perceived consensus of opinion.* We obtained a measure of the accuracy with which member attitudes were perceived by having each *S* identify those participants whom they thought held opinions similar to and divergent from their own.

The results showed that low Es were generally more accurate in identifying the attitudes held by other members. In addition, there were significant differences in the degree of accuracy for both high and low Es in the various structure conditions. Interestingly, the source of these differences was the same for both attitude groups. That is, *S*s were most sensitive to attitudinal differences when judging from a minority position and were least sensitive in the balanced groups.

DISCUSSION

The results of the experiment may be examined with respect to our initial aims as well as to certain hypotheses suggested earlier by Festinger (1950).

Festinger predicts, for example, that increased pressure to communicate to others in a group occurs with increases in the perceived opinion discrepancy among members. Our data enabled us to determine the extent to which the group members accurately perceived one another's ethnocentric bias as well as the direction, rate, and spontaneity of their discussion behavior. Since the low Es were more accurate than the high Es in identifying the attitudes held by others in their group, it follows that the low Es should have been more impelled to communicate than the high Es, who were less cognizant of intragroup discrepancies in attitude. Our results support this prediction, since the low Es displayed higher rates of response than the high Es. The data show that this difference in response rates between high and low Es derived mainly from groups in which the low Es were in a minority, and it was precisely these groups that the low Es were most sensitive to attitudinal differences among the group members. Of course, the generally greater discussion activity of the low Es could have been due in part to the fact that the communication supported their point of view.

An analysis of the low E response rate shows an increased tendency for these *S*s, when in the minority position, to interact with both low Es and high Es. It might have been predicted that the low Es, with film support of their views, would have directed the preponderance of their comments toward the "deviant" high Es. This, in fact, was the case, although they also increased the number of remarks directed toward those holding similar views. Thus, as their number decreased, and as they became more aware of the opinions held by the other members of the group, the low Es increased their rate of communication to both the opposition and to those who agreed with them. Apparently, communication behavior is not designed solely to influence those who disagree with one's own views, but also aims at providing and receiving support from those holding similar views.

We obtained support for Festinger's hypothesis that increased pressure to communicate with others occurs with increases in the cohesiveness of the group. Low Es, who had the highest rates of response, were more attracted to their groups than high Es. Attraction to the group for the low Es, however, was not related to group structure. In the case of the high Es, the greatest attraction to the group was reported by those in the majority position; and it was in these groups that they communicated more with the low Es than they did with each other. Attraction to the group, then, to extend Festinger's suggestion, is apparently an important predictor not only of the amount, but of direction of communication.

On the sociometric item, the high Es divided their choices about equally between low Es and other high Es, regardless of group structure. The low Es, however, showed significantly greater preference for one another when they were in either a majority or a minority position. We suspect that this may be a reflection of the more active discussions held under these two conditions, which gave the participants a greater opportunity to assess one another's attitudes. In fact, it was in these groups that the perceived consensus judgments by low Es were most accurate. Having assessed fairly accurately the ethnocentric dispositions of other members, the low Es preferred those individuals whom they perceived as more like themselves. The high Es were not as perceptive of differences in group structure and, therefore, could not be expected to assign their sociometric choices on other than a 50–50 basis.

Finally, we may note the interesting behavior displayed by the balanced groups, which were composed equally of high and low Es. Under these conditions, spontaneity, rate of response, and number of opposition-directed comments were generally lower than in any of the other structures. It may have been predicted that maximum opinion discrepancy would be associated with greater discussion activity. Our findings also indicate that the members of the balanced groups were least attracted to their groups, were least accurate in judgments of each other's attitudes, and were less likely to feel compatible with those individuals holding opinions similar to their own. These results suggest that where opinion equilibrium exists in a group, discussion activity is suppressed, with a resultant loss in accuracy of interpersonal perception and group cohesiveness. One is tempted, in view of the previous finding by McGinnies and Altman (1959) that middle E groups were less active in group discussion than either low Es or high Es, to speculate that something analogous to an algebraic summation occurs when equal numbers of low Es and high Es form a group. A balancing of attitudes seems to produce discussion behavior similar to that in neutral groups.

SUMMARY

The purpose of this study was to investigate the effects of various degrees of minority-majority conditions on member behavior during the course of group discussions and to ascertain some of the relationships between various aspects of interpersonal perceptions and discussion behavior.

On the basis of their scores on a modified version of the California E scale, high school students were randomly assigned to one of five discussion conditions representing various degrees of group composition. Two group structures had homogeneous membership with respect to ethnocentric attitudes, that is, all had either high or low scores on the California E scale. The remaining three group structures were heterogeneous with respect to E scale scores and contained either low Es in the majority, high Es in the majority, or equal numbers of each.

Each six-man group viewed and discussed a film dealing with the general problem of ethnic minorities and then completed a questionnaire measuring various aspects of interpersonal perceptions. Discussion behavior was measured by a series of indices coded during the course of group meetings.

With one exception, the results showed no clear-cut and simple group structure effects, but indicated an interaction between group composition and interpersonal perceptions as determinants of discussion behavior. The major exception occurred for groups composed of equal numbers of high and low Es. In these groups, spontaneity, rate of response, and number of opposition-directed communications were generally lower than in any of the other structures. Members of these groups also were least attracted to their groups, were least accurate in their perceptions of opinions held by other members, and were less likely to choose sociometrically others in the group holding views congruent to their own. These findings suggest that opinion "equilibrium" results in a suppression of discussion activity and concomitant pressures to influence members holding divergent points of view.

The results of the study indicated an intimate linkage between interpersonal perceptions and discussion behavior. Those members who were most accurate at identifying the opinions held by others in the group played a larger role in discussions than those who were not so accurate. This effect was most pronounced in the heterogeneous structures, where there were divergencies of opinion. In addition, there was a tendency for discussion pace and opposition-directed communication to be positively related to valuation of the group.

The findings lend some support, as well as some qualifications, to Festinger's conclusions regarding influence effects in small groups.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

DICKENS, M. A statistical formula to quantify the "spread-of-participation" in group discussion. *Speech Monogr.*, 1955, **1,** 28–31.

FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57,** 271–292.

FESTINGER, L., BACK, K., SCHACTER, S., KELLEY, H., & THIBAUT, J. *Theory and experiment in social communication*. Ann Arbor, Mich.: Institute for Social Research, 1950.

FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46,** 92–99.

GERARD, H. B. The effect of different dimensions of disagreement on the communication process in small groups. *Hum. Relat.*, 1954, **3,** 313–327.

MCGINNIES, E. M., & ALTMAN I. Discussion as a function of attitudes and content of a persuasive communication. *J. appl. Psychol.*, 1959, **43,** 53–59.

MITNICK, L. L., & MCGINNIES, E. M. Influencing ethnocentrism in small discussion groups through a film communication. *J. abnorm. soc. Psychol.*, 1958, **56,** 82–90.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, New York, 1935, **27,** No. 187.

(Received June 2, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

CONDITIONING AND GENERALIZATION OF GSR AS A FUNCTION OF AWARENESS[1]

B. B. CHATTERJEE AND CHARLES W. ERIKSEN
University of Illinois

In 1937, Diven reported an experiment purporting to demonstrate the conditioning of anxiety reactions in human subjects without their awareness of the conditioned stimulus. His procedure involved a list of stimulus words in which the word "Barn" occurred six times. The subjects (*S*s) were required to chain associate to each of these stimulus words, and following their chain associations to the word Barn *S*s were given a strong electrical shock. Diven reported that *S*s, although unable to verbalize the contingent relationship between barn and shock, gave significantly greater GSRs to the word Barn than did a control group of *S*s who had not been shocked. Further, he reported that the conditioned GSR showed semantic generalization to other rural words that had been contained in the stimulus list. Haggard (1943) in a somewhat similar experiment obtained comparable results.

For obvious reasons, these two experiments have played an important role in theorizing about psychodynamics. White (1948) among others has leaned extensively on the Diven experiment in his theory of anxiety and the learning of neuroses.

There are a number of methodological flaws in both the Diven and Haggard experiments that make the conclusion of unconscious conditioning of anxiety at best very tenuous. Several of these defects have been pointed out by Lacey and Smith (1954) who repeated the Diven experiment with an improved design. They used a somewhat different word list and two experimental groups. *S*s in the Cow-shock group were given shocks following their associations to the word Cow while *S*s in the Paper-shock group were shocked following their associations to the word Paper. Heart rate rather than GSR was used as the conditioned response, and the anticipatory changes in heart rate that occurred during the *S*'s associations were analyzed instead of the responses following association, as was the case in the Diven and Haggard studies. These changes in method, however, appear to have had little effect since Lacey and Smith conclude that not only did both groups of *S*s show conditioning of anxiety reactions without awareness but also generalized the conditioned response to semantically similar words.

In spite of their changes in design, the Lacey and Smith study still lacks some crucial controls that would make the interpretation of unconscious autonomic conditioning unequivocal. Eriksen (1958) has pointed out that their method of determining awareness in *S*s and their method of computing conditioning are such as to permit a spurious precision in the autonomic conditioning relative to the verbal awareness.

They determined awareness by intensively questioning *S*s as to whether they had learned when to expect shock. *S*s classified as unaware were those who were unable to verbalize the specific contingent relationship between shock and the word Cow (Paper). This method of questioning, however, may not elicit all the information available to the *S*'s verbal report. It is quite possible that the *S*s, if asked, could have given several or more words which they were quite confident were not shocked. It seems reasonable to assume that the *S*'s change of heart rate to a stimulus word which he is confident will not be followed by shock will be less than to those words of which he is still unsure (potentially dangerous). It would also seem reasonable to expect the actual shock word in general to be among those words that the *S* was uncertain of rather than among the words he was rather confident were not followed by shock, while a semantically dissimilar word would more likely be in the latter category.

The evidence that previous studies have found on GSR generalization to words in the same semantic classification might actually represent not generalization so much as a beginning on the part of the *S* to learn to discriminate the general area where shock occurs. Thus, *S*s in the Cow-shock group would tend to classify rural words in the uncertain category and nonrural words including paper in the confident not shock category. The reverse would occur for the paper shock *S*s. In view of this possibility it becomes apparent that the control measure used to determine whether conditioning is occurring becomes quite crucial.

---

[1] This research was supported by Mental Health Grant M-1206, National Institutes of Health, Public Health Service.

In the study by Lacey and Smith, conditioning was computed by taking each S's heart rate score to cow and subtracting from it his heart rate score to the word paper. It should be readily apparent that this method of computing conditioning is apt to lead to a spurious appearance of precise discrimination by the heart rate response which in fact may not exist. To show that conditioning of heart rate is more specific than the stage of the *S*'s verbal awareness, an analysis is required that demonstrates whether or not the *S* shows greater autonomic conditioning to the actual shock word than he does to the rest of the words concerning which he verbalizes uncertainty as to whether they were followed by shock.

The experiment reported below replicates the Lacey and Smith study in its essential details with the exception that *S*'s verbal awareness was explored more extensively. The *S*'s verbalizable expectancies concerning shock were obtained for each stimulus word in order to control for the possible artifact described above. In addition, some of the characteristics of the experimental situation that were related to the *S*s' expectancies were explored.

METHOD

The procedure was administered to each *S* individually. *S*s were comfortably seated with the experimental apparatus screened from their view and were instructed that this was an experiment on the effects of tension on mental associations. They were further told that a list of stimulus words would be presented to them one at a time by means of a tape recorder, and upon the occurrence of a stimulus word, they were to chain associate to this word until the stop signal occurred. The process of chain association was explained to them, and then they were given practice in chain association to three nonexperimental words to insure that they had adequately comprehended the directions.

Next, the shock electrodes were connected to the right forearm of the *S*, 2 in. above the wrist with an interelectrode distance of 1 in. The electrodes were fixed firmly into position by masking tape. The *S* was told that the shock was an integral part of the experiment designed to produce tension on his part and would occur several times throughout the word list. At this point, the *S* was given several sample shocks in order to adjust the voltage level of the shock to the point where the *S* reported it as very painful. The GSR was recorded from the dorsal and ventral surface of the middle finger of the left hand through silver electrodes. The finger was cleaned with alcohol and treated with electrode jelly. GSR records were recorded continuously throughout the experimental session.

Upon completion of the instructions and after a short period of adaptation, the experimental session began. The tape recorder presented the stimulus words in the same order as that employed by Lacey and Smith. The words had been recorded by a person trained in speech to minimize distortion or confusability in pronounciations. Following the occurrence of the stimulus word, 15 sec. were allowed for the *S*'s chain associations before the word stop occurred. Fifteen seconds then elapsed between the stop signal and the occurrence of the next stimulus word. The 40-item stimulus list consisted of the words, Cow, Paper, Easy, and Harvest six times each, eight other rural words and eight nonrural words (common adjectives and nouns) all distributed in temporally balanced order in the list.

During the conditioning session (one presentation of the 40-item list), one group of *S*s (Cow-shock) was administered a painful shock coinciding with the stop signal following each presentation of the word Cow. The other group of *S*s (Paper-shock group) were shocked similarly following the word Paper.

Upon completion of the conditioning phase, the following questions were asked and the *S*'s answers recorded verbatim:

1. Did you know when you were going to get shocked?
2. Please explain freely what you think about the shocks in this experiment.
3. Please try to explain to me what you think or guess is the scheme behind the administration of the shock. Try to guess in any manner you like—you can guess as much as you like, all I'm interested in is what you think about or make out concerning shock.
4. Let me read the list of stimulus words and let us see whether you can recognize any particular word or words which might have been followed by shock. I want to see whether you can distinguish between words which might have been followed by shock and those which might not have been followed by shock. You can say No, Yes, Uncertain or whatever you think concerning the probability of the particular word having been followed by shock. If you're not sure, guess.

An extinction phase followed the inquiry. It was identical to the conditioning session with the exception that no shocks were administered although the electrodes were left undisturbed and the *S*s were not informed that there would be no further shocks.

Apparatus. Electric shock was administered by means of a Harvard inductorium using a faradic stimulation. The primary coil was supplied from a 3-v. dc source. The voltage of the secondary was adjusted for each S to a level that was reported very painful.

The GSRs were continuously recorded with a modified Deceptograph supplied by C. H. Stoelting and Company. The modification was such as to permit the continuous reading of the S's basal resistance level in ohms.

Subjects. The 22 Ss consisted of 4 men and 18 women who were volunteers among graduate students in the College of Education at the University of Illinois. They were randomly assigned to the Cow-shock and Paper-shock groups. None was familiar with the original Diven experiment nor the experiments of Haggard and Lacey and Smith. The Cow-shock group contained 12 Ss and the Paper-shock group 10.

RESULTS

On the basis of the Ss' verbalizations during the inquiry, they were classified into aware and unaware Ss. Three Ss in the Cow-shock group and one in the Paper-shock group were able to correctly verbalize the contingent relation between shock and the correct stimulus word. The remaining 18 Ss were classified as unaware. Their verbalizations were vague concerning the stimulus word followed by shock, although some verbalized incorrectly but with confidence that particular words had been followed by shock one or more times during the conditioning session. The four aware Ss have been excluded from the following analyses.

The first step in the analysis was to determine the extent to which the present data duplicated the findings of Lacey and Smith. Complete GSR records were available on seven unaware Ss in each of the two shock groups.[2] The maximum GSR deflection in micromohs of conductance change that occurred during the 15-sec. association interval was determined for each of the 40 stimulus words for each S. The list of stimulus words was then divided into four blocks of ten consecutive words each and the mean and SD of the GSR deflections for each S for each of these four blocks was computed. GSRs to the different words were then converted into z scores based upon the mean and SD of the 10-word block in which they occurred. These transformed GSR z scores were used in all the subsequent analysis.

To render our data comparable with those of Lacey and Smith, their method of determining conditioning was first computed. In this method, the transformed GSR z score to paper (in their study, heart-rate) is subtracted from the response to Cow for each S. Since words in corresponding positions are taken for comparison, any temporal trend is controlled. The technique of moving averages by taking the mean of consecutive blocks of three trials is used to smooth out irregularities in the conditioning curves for the two shock groups.

The results are graphed in Fig. 1. In terms of this measure of "conditioning" the Paper-shock group gives evidence similar to that obtained by Lacey and Smith with the maximum GSR occurring on the sixth block of trials, following which extinction takes place. For the Cow-shock group, however, there appears to be no consistent evidence of conditioning until approximately the seventh trial which is in the extinction period and after the pause for questioning. By the tenth trial, however, there is some evidence of extinction in this group.

Lacey and Smith evaluated the significance of these trends by the nonparametric one sample runs test applied to the response differences block by block. Applying this test to our data, we find that only in the seventh block was a difference between the two groups sig-

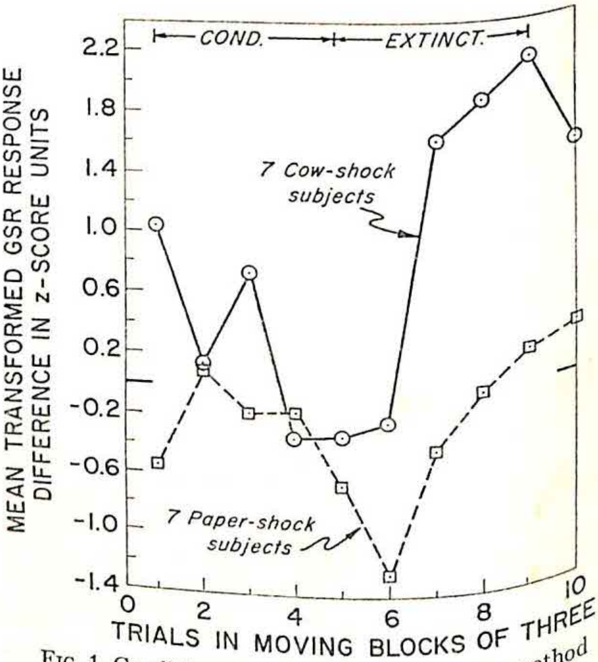


FIG. 1. Conditioning curves computed by the method of Lacey and Smith (1954).

---

[2] Due to recording difficulties, the GSR records for two Ss in each group were not complete. These four Ss were excluded from the analyses of the GSR data.

TABLE 1
MEAN TRANSFORMED GSR READINGS FOR THE SHOCK GROUPS FOR DIFFERENT GROUPINGS OF WORDS DURING THE CONDITIONING SESSION

Shock-group	Shock Word	Repeated Similar Words	Repeated Nonsimilar Words	Nonrepeated Similar Words	Nonrepeated Nonsimilar Words
Cow-shock group	0.38	−0.044	−0.42	1.81	−0.93
Paper-shock group	1.82	1.37	−1.33	2.29	−0.58

nificant at the .05 level (one-tailed). However, summing for each *S* through the six conditioning trials and testing the mean difference between the Cow-shock and Paper-shock group resulted in a *t* value of 1.30, which with 12 degrees of freedom is not statistically significant. Summing through eight trials (taking the six conditioning trials and the first two extinction trials together) and then testing the mean difference for the two shock groups yields a *t* of 2.83, which is significant beyond the .01 level (one-tailed). A similar *t* test applied to the sum of the six extinction trials resulted in a value of 2.31, which again with 12 degrees of freedom is significant beyond the .03 level (one-tailed).

Evidence of conditioning within each of the two groups during the conditioning phase was evaluated by a three-way analysis of variance (trials, Cow and Paper, *S*s) of the *z* scores to Cow and Paper. A separate analysis was performed for each group. In these analyses conditioning to the shock word would be reflected in a significant interaction between trials and the Cow and Paper words.

Only in the Paper-shock group was there evidence of conditioning with the interaction term significant at approximately the .05 level ($F = 2.50$ with 5 and 30 *df*). The main effects did not approach significance in either group and the absence of significant interaction in the Cow-shock group provides no evidence of conditioning for these *S*s.

Lacey and Smith also found that the conditioned responses tended to generalize in the respective groups to words of a similar semantic classification. Thus, the Paper-shock group showed a greater response to nonrural words than to rural words, whereas the Cow-shock group gave greater responses to the rural words than to the nonrural words. The present data were examined for evidence of this type of generalization. In Table 1, the mean transformed GSR readings for the two experimental groups during the conditioning session are shown as a function of different groupings for the words. As is seen, the mean GSR for the nonrepeated similar words in both groups is higher than that for the actual shock words, a finding which is in conformity with that of Lacey and Smith who also noted that the generalized response magnitude tended to be higher than the response to the actual conditioned stimulus.

The significance of the differences observed in Table 1 was evaluated by a modified three-way classification analysis of variance (groups, word classification, and individuals). The word classification was broken down into shock words, repeated similar words, repeated nonsimilar words, nonrepeated similar words, and non-repeated nonsimilar words. The results of this analysis failed to show any significant effects, with the *F* ratio for between word classifications only 1.54 (4 and 48 *df*, $p > .05$).

Since, as will be seen below, repetition of words in the list has a confounding effect upon any measure of conditioning or awareness, a more meaningful test for generalization is in terms of the nonrepeated rural and nonrepeated non-rural words in the list. Again a modified three-way classification analysis of variance similar to the above was performed but in this case only two word classifications were used, nonrepeated rural and nonrepeated nonrural words. None of the main effects in this analysis was significant, but the interaction between the two shock groups and rural-nonrural word classification was significant beyond the .05 level ($F = 5.87$ with 1 and 12 *df*). This significant interaction reflects the trend that can be seen in the above table where it is found that the Paper-shock group shows greater GSR to nonrural words while the reverse effect is true for the Cow-shock group. Similar analysis of variance performed upon the extinction trials failed to yield any significant effects.

The data analyzed so far are in general agreement with the results presented by Lacey and Smith. For the Paper-shock group, at least, they could be interpreted in terms of rather

TABLE 2

LEVEL OF CERTAINTY OF SHOCK SCALE AND CORRESPONDING SCALE SCORES

Scale Score	Scale Category	Verbal Responses included in Each Scale Category
6	Yes	Yes
5	Once or twice	Maybe once or twice, Once, Once or twice
4	Maybe	Maybe, Perhaps, Possibly, Could be, Probably, I think so, I believe so
3	Uncertain	Uncertain, Not sure
2	I don't know	I don't know, I don't remember, Never heard, Didn't hear, Can't see any connection
1	I don't think so	I don't think so
0	No	No

specific autonomic conditioning in the absence of verbal awareness. However, as was previously pointed out, this evidence for specificity of conditioning might very well be an artifact of the method of computing or analyzing conditioning. The analysis that has been conducted so far has not determined verbal awareness in the *S*s in a manner that is commensurate with the analysis of autonomic discrimination or conditioning. While all of the *S*s contributing to the data were unaware in the sense that they could not verbally report that shock followed Cow (Paper), they nonetheless had various expectancies concerning the relations between certain words and shock. The existence of these expectancies was revealed by the more detailed assessment of awareness that we employed in this study.

As will be recalled, we presented each of the 20 different stimulus words to the *S* and asked him to state whether or not these particular words had been followed by shock. It was possible to sort their replies on a seven-point scale defining the expectancy level of the *S* as to whether a word had been followed by shock. Table 2 shows how the Ss' replies were classified in terms of the scale scores.

For the *S*s in the Paper-shock group, the average number of words per *S* receiving a rating of three or higher was 5.4 (range 3–8). For the *S*s in the Cow-shock group, the mean was 6.8 (range 5–12). It was further found that five of the seven *S*s in the Paper-shock group had given a value of three or higher to the word paper, and four of the seven *S*s in the Cow-shock group had given a value of three or higher to the word Cow. The words receiving scale values of three or higher may be considered to be perceived by the *S*s as potentially dangerous relative to the words having scale values that are lower than three.

Our next step was to determine whether the GSR response was greater to the actual shock word than it was to the other words that *S* perceives as potentially dangerous. For each *S*, the GSRs in z score units during the conditioning trials were averaged for all words to which he had given a rating of three or higher, excluding the actual shock word. The difference between this mean GSR to the words perceived as dangerous and the GSR to the actual shock word was computed for each *S*, and the mean of the differences by experimental group was evaluated with a t test to see whether it differed significantly from zero. For the Paper-shock group, the mean of the z scores for the shock word was .18 as was compared with a mean of .26 for the words with scale values of three or more. The mean difference of .08 although not significantly different from zero ($t = .09$) is actually in the opposite direction from what would be expected if the *S* were unconsciously discriminating the true shock word from the other words perceived as dangerous. For the Cow-shock group, the comparable means were .06 and $-.16$. Again the mean difference (.22) although now in the direction for unconscious conditioning, is not significantly different from zero ($t = 1.54$, $p > .05$). It would appear therefore, that in the present experiment the autonomic conditioning manifested by the *S*s is no more specific or precise than their verbally stated awareness.

It is possible now to show not only that the evidence given by the Paper-shock group for autonomic conditioning without awareness is artifactual, but also why the Cow-shock group did not show such evidence. Examination of the *S*'s expectancy of shock for the different words shows that only two of the seven *S*s in the Paper-shock group report having been shocked to the word *Cow* while five of the seven *S*s in the Cow-shock group report having been shocked to *Paper*. Since the GSR tends to be greater to words which *S*s report as having been followed by shock, assessing conditioning for the *S*s in the Paper-shock group by sub-

TABLE 3

KENDALL'S COEFFICIENT OF CONCORDANCE, W, OBTAINED FROM DISTRIBUTION OF RANKED LEVEL OF CERTAINTY OF SHOCK FOR DIFFERENT GROUPINGS OF SUBJECTS AND STIMULUS WORDS

Grouping	No. of Ss.	No. of Words	Kendall's Coeff. (W)	Significance (p)
Cow-shock group: Rural words only	9	10	0.127	.30
Cow-shock group: Nonrural words only	9	10	0.160	.15
Cow-shock group: All words together	9	20	0.129	.20
Paper-shock group: Rural words only	9	10	0.255	.01
Paper-shock group: Nonrural words only	9	10	0.468	.01
Paper-shock group: All words together	9	20	0.360	.01

tracting the autonomic response to paper from the autonomic response to Cow would give the general results shown in Fig. 1.[3] Also, the fact that a high number of Cow-shock *S*s report having been shocked to Paper would lead us to expect negative results from this type of subtractive computation.

It is informative to analyze the expectancies of the *S*s further in order to identify features and characteristics of the experimental arrangement that lead *S*s to identify certain of the stimulus words as being associated with shock. First, we examined the consistency among *S*s within experimental groups in the assignment of their expectancies to different stimulus words. For this purpose, Kendall's coefficient of concordance (W) was employed. Table 3 shows the Ws for rural words, nonrural words, and all words together between *S*s in each of the two experimental groups.[4]

As can be seen from the table, the *S*s in the Cow-shock group did not show a significant amount of consistency in their expectancies for the different words together or when broken down into separate categories. The Paper-shock group on the other hand showed significant consistency in all three analyses. Thus, only in the Paper-shock group is there evidence that the administration of shock had a systematic effect upon the hypotheses the *S*s construct concerning shock.

Despite lack of consistency among the Cow-shock *S*s there is actually considerable consistency in the assignment of certainty to words between the two shock groups. If the average expectancy for each of the 20 different stimulus words is computed separately for the two experimental groups, the rank order correlation of these expectancies between groups for the 20 words is .62 ($p < .05$). This finding would suggest that there are some elements in the experimental arrangement that biased the *S*s' ideas concerning the occurrence of shock.

To identify some of these factors, the stimulus words were subdivided into five categories: the shock word; repeated similar words; repeated dissimilar words; nonrepeated similar words; and nonrepeated dissimilar words. A modified three-way classification analysis of variance (groups, word classification and individuals) were performed on the expectancy scores for the different words. This analysis revealed a significant variation between word classification ($F = 6.81$ with 4 and 64 *df*, $p < .01$), but there were no significant differences between the two groups nor was the interaction between groups by word classification significant. Inspection of the cell means from this analysis suggests that the significant effect due to word classification is attributable to the tendency for words that have been repeated six times in the list to have higher shock expectancy than words which have occurred only once, and, further, that words belonging to different semantic categories are also being discriminated with regard to expectancy. Apparently, the *S*s have begun to differentiate at the verbal level the experimental situation in terms of two relevant characteristics: the repetition of certain stimulus words and semantic classification.

---

[3] This method of assessing conditioning for the two *S*s in the Paper-shock group who reported having been shocked to Cow shows little or no evidence of conditioning. This result is to be expected in view of the fact that they had also included Cow as a potentially dangerous word in their verbal reports.

[4] For these analyses, N in each group is nine since the expectancy data were obtained for two *S*s in each group on which complete GSR data was not available.

DISCUSSION

Lacey and Smith (1954) in their penetrating analyses of the Diven (1937) and the Haggard (1943) experiments have quite accurately pointed out that neither of the experiments contained the necessary controls to support a conclusion of "unconscious conditioning." Diven compared the GSR of the experimental *S*s to the shocked word with that of a control group who had received no shocks at all. As Lacey and Smith (1954) point out, this comparison is invalid in the absence of a demonstration that the experimental group's reactivity to all words was not greater than the control group's. This criticism also applied to Haggard's study. And as was seen above, while Lacey and Smith use a more sophisticated and appropriate statistical test of conditioning, they still have not excluded the likelihood that their evidence of "unconscious conditioning of anxiety" is artifactual.

The methodological flaw in the Lacey and Smith study is the same one that has plagued a number of attempts to demonstrate learning without awareness. Adams (1957) has termed it the problem of correlated hypotheses and has shown how failure to consider this problem has invalidated many of the experiments in this area.

The problem arises because these experiments focus on and record only a limited aspect of the *S*s' behavior. When the *S* is examined for awareness, he is classified as unaware if he is unable to repeat the verbalization the *E* has formulated. But his over-all behavior in the experiment may be in perfect correspondence with his verbalizations. What the experimenter overlooks is that the change or modification in the observed response could actually be correlated with a number of different hypotheses. Convincing research in learning without awareness must of necessity thoroughly explore the *S*s verbalizable hypotheses and expectations and only when it is shown that these hypotheses are unable to account for the observed change in the response is a conclusion of learning without awareness warranted.

At present we are left with no convincing evidence that specific autonomic conditioning occurs in the absence of verbal awareness. Razran (1949) has reported salivary conditioning in human *S*s without their awareness, but as Lacey and Smith point out, Razran's description of his method of determining awareness is so sketchy as to leave serious doubts as to whether he successfully avoided the pitfalls that have been discussed.

We would suggest that the future investigations in this area would do well to abandon the method originated by Diven (1937). Our results suggest that there are so many uncontrolled and confounded sources of variance in this design as to render virtually impossible an answer to the basic question of conditioning without awareness. Both the *S*'s expectancies of shock as well as his GSRs were a function of the frequency of repetition of certain words in the list. The introduction of new words toward the end of the list also seemed to influence the GSR, perhaps as a result of their novelty or unexpectedness. In addition, the 15-sec. CS–US interval is quite long for conditioning to occur, since the optimum interval is generally considered to be much shorter.

Generalization of the autonomic response would more appropriately be measured by the introduction of new words after the conditioning trials have taken place. The present design results in a certain amount of masking of what generalization effects there are due to the fact that the *S* is presented the generalization stimuli interspersed with the conditioning trials and thus is learning a condiioned discrimination.

There is one other aspect of our data that deserves brief comment. This is the increase in GSR to the conditioned stimulus in the Cow-shock group during the extinction trials, (Fig. 1), an effect which is particularly interesting since the *S*s in this group gave no evidence of conditioning during the six conditioning trials. It is impossible to evaluate the meaning of this effect since, unfortunately, we did not requestion the *S*s following the extinction phase of the experiment. What changes may have occurred in their expectancies concerning shock as a result of the inquiry period can only be surmised.

SUMMARY

This study was essentially a replication of the Lacey and Smith (1954) experiment on the unconscious conditioning of autonomic responses. When their methods of data analysis were used, their results in general were con-

firmed. However, due to a more extensive and careful investigation of the *S*s' awareness in the present study, it was possible to show that the conditioning of the autonomic response (GSR) was no more specific than the *S*s' verbalizations. It was also shown that the method employed by Lacey and Smith to compute conditioning gives rise to a spurious conclusion of precise autonomic conditioning without verbal awareness. The limitations of their basic design for investigation of the problem of unconscious conditioning were pointed out.

REFERENCES

Adams, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

Diven, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.

Eriksen, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1958.

Haggard, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomous reactions during and following induced stress. *J. exp. Psychol.*, 1943, **33**, 257–284.

Lacey, J. L., & Smith, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.

Razran, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.

White, R. W. *The abnormal personality.* New York: Ronald, 1948.

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

THE EFFECTS OF SUBLIMINAL AND SUPRALIMINAL SUGGESTION ON VERBAL PRODUCTIVITY[1]

MARVIN ZUCKERMAN

Institute of Psychiatric Research, Indiana University Medical Center, Indianapolis

INVESTIGATIONS of "subliminal" influences on behavior are not new in psychology; studies in this area date back to the 19th century. However, judging from recent reviews of the literature by Adams (1957), McConnell, Cutter, and McNeil (1958), Eriksen (1958), and Goldiamond (1958), a resurgence of interest in the area is underway.

Most of the studies reviewed by Adams (1957) require a subject to make a stimulus discrimination. In one group of these experiments, the subject knows the nature of the cue but his discrimination is better than chance at stimulus intensities at which he thinks he is guessing. In another group of experiments the subject does not know that a cue to discrimination is being given. Only 9 of the 57 studies in Adam's review do *not* involve a stimulus discrimination task. Seven of these are essentially operant conditioning studies in which the subject does not realize he is being conditioned because he is unaware of either the reinforcer or the class of responses being reinforced, or of both. The use of subliminal suggestion in movie and television advertising seems to belong with this latter group, for the subject's operant behavior is supposedly brought under control by suggestive cues of which he is not aware. The present study represents an attempt to construct an analogue of this type of situation. The question asked in the first experiment was: "Can one influence frequency of a graphomotor response by suggestions administered visually at subliminal intensities?"

EXPERIMENT I

The experimenter tried to control the *S*'s verbal productivity (the number of words written in describing pictures) with two subliminal suggestions: WRITE MORE and DON'T WRITE. A control group of *S*s described the pictures without subliminal suggestion. Each of these suggestions was typed in capital letters on a paper-roller fitted into a tachistoscope. Three TAT cards—4, 6BM, and 7GF—were used as the pictures to be described in writing. These pictures, which each portray two persons with one turned away from the other, were rotated over the conditions in the experiment so that each picture appeared an equal number of times in each condition.

Subjects

The 36 *S*s consisted of 33 student nurses and three secretaries. Their average age was 20.0 ($SD = 1.3$). The student nurse *S*s were paid volunteers.

Apparatus

The apparatus used was the Harvard Tachistoscope, a modification of the Dodge Mirror Tachistoscope. This model, manufactured by Ralph Gerbrands Co., has dual timers permitting control of two stimulus fields. Its fluorescent lamps may be turned on and off instantly. Dual mechanical timers provide exposure intervals of .01 to 1 sec. for two sets of lamps, each set controlling a field of exposure. The duration of the overlap or interval in the presentation of the fields is controlled by a third timing mechanism. Single pieces of Atlantic bond paper (20 wt.) were used as filters on the two lamps controlling the second exposure field.

Procedure

The *S*s were told that they would be shown three pictures in the tachistoscope and that each picture would be shown 10 times, each time for a slightly longer interval than the preceding time. After each presentation, they were to write on a pad everything they saw in the picture, including everything that was happening in the picture. They were told that it was permissible to repeat what they had written on previous presentations of the same picture but they were told not to use ditto marks or words like "same as before."

Before starting the experiment the word "SAMPLE" was presented for .02 sec. against a blank, lit background field. The *S* was asked if she had seen either a word or letters presented. If (as was the case with two *S*s) she did report a word or letters, one additional bond paper filter was inserted in front of each of the two lamps controlling the second exposure field, which contained the word; to insure that the experimental words would be subliminal. All but two *S*s reported seeing merely a "flash," a "blur," or, in a few cases, "lines."

---

[1] James A. Norton, Department of Mathematics and Statistics, Purdue University, was the statistical consultant for this study. He devised the modifications of standard designs and tests of significance used in the analyses of the data.

Condition I was the same for both experimental and control *Ss*. The first exposure field, containing a TAT picture, was exposed ten times for successive durations beginning at .05 sec. on the first trial and lengthening .05 sec. on each successive trial, reaching .5 sec. on the 10th trial (ostensibly, the experiment was to test the effect of different durations of presentation on perceptual acuity). The second exposure field was blacked out except for an exposure slot in which the stimuli were to appear during the experimental conditions. During Condition I this slot was blank, resulting in a flash of .02 sec. duration overlapping with the last .02 sec. of the presentation of the TAT picture and superimposed on the picture. After each trial the *S* was allowed to write her description of the picture. She was instructed to say "ready" whenever she finished writing. No time limits were placed on the *S*. However, because of the close scheduling of *Ss* it was necessary to discard *Ss* who took longer than half an hour to complete the ten trials of Condition I. Three *Ss* were discarded because of this limitation. As soon as the *S* signalled that she was ready for the next presentation, the examiner gave a ready signal and pressed the timer button.

In Condition II, a second TAT card was presented. *Ss* followed the same procedure as in Condition I. The *Ss* in the control group again had a blank in the second exposure field. However, the *Ss* in the experimental group were given the stimulus words "WRITE MORE" in the second exposure field on each of the ten trials. This stimulus overlapped for .02 sec. with the last .02 sec. of the picture being presented in the first field. It appeared approximately in the middle of the picture.

In Condition III a third TAT picture was presented with the *Ss* following the same procedure. The *Ss* in the control group again had a blank in the second exposure field. The *Ss* in the experimental group were given the stimulus words "DON'T WRITE" in the second exposure field, again overlapping for .02 sec. with the end of the exposure period of the picture.

After the last trial on the last condition, a blank card was inserted in the first exposure field which now served as a lit background field. The stimulus words from the last condition were presented at the .02 sec. setting and the *S* was asked to describe what she saw. The stimulus presentation was lengthened .01 sec. on each successive trial until the *S* could correctly read it. This threshold was obtained in order to determine how close to threshold the *S* was functioning during the last condition of the experiment. The mean threshold for the 18 *Ss* in the experimental group was .08 sec. (*SD* = .04). However, it was discovered in the second experiment to be reported that it is much more difficult for *Ss* to recognize words presented at the end of a primary stimulus to which they are attending, than it is to recognize the stimulus against a blank background. It is likely that the actual threshold for the words against the background of a complex picture was much higher than the threshold obtained against a blank background. However, on the assumption that this measure might have some correlation with the real threshold, it was examined in relation to the magnitude of the suggestion effect.

TABLE 1

EXPERIMENTAL DESIGN

Group and Card Order	*N*	Conditions I	Conditions II	Conditions III
Control				
ABC	6	.02-sec. blank	.02-sec. blank	.02-sec. blank
BCA	6			
CAB	6			
Total	18			
Experimental				
ABC	6	.02-sec. blank	.02-sec. words ("Write More")	.02-sec. words ("Don't Write")
BCA	6			
CAB	6			
Total	18			
Total *Ss*	36			

After the threshold was obtained, *Ss* were carefully questioned as to whether they saw any words being presented with the pictures. Only one *S* in these two groups reported seeing the stimulus words and she was discarded as a *S* in the experiment.

Ss were urged not to tell any details of the experiment to other nurses. Actually almost all of them were still unaware of the nature of the experiment. Most of them seemed surprised when they were asked if they noticed words presented with the pictures. New *Ss* were questioned as to what they had heard about the experiment and as far as could be determined, none of them had heard that words were presented along with the pictures.

Design

The design is presented schematically in Table 1. The design was one of replicated 3 × 3 latin squares. The complete analysis of this design is an extension of the analysis presented by Edwards (1950, p. 326).

Results

The "Residual within Individuals" sum of squares in the design is made up of six parts: a Subjects × Conditions interaction (with 10 *df*) from each of six Group × Order of Presentation combinations. These six contributions were computed separately and tested for homogeneity with the Bartlett test, which yielded a chi square of 15.09 (significant at the .05 level). Since the variance was heterogeneous a logarithmic transformation was employed. Bartlett's test was applied

TABLE 2

ANALYSIS OF VARIANCE

Source of Variance	*df*	Mean Square	*F*
Between Groups	1	.298	2.54
Between Orders	2	.203	1.73
Groups × Orders	2	.023	.20
Between *S*s within Groups × Orders	30	.117	
Between Conditions (avg. both groups)	2	.101	12.64**
Conditions × Groups	2	.027	3.44*
Between Cards	2	.001	.16
Cards × Groups	2	.013	1.66
Latin Sq. Residual	2	.006	.75
Latin Sq. Residual × Groups	2	.030	3.81*
Residual within *S*s	60	.008	
Total	107		

* $p = .05$.

** $p = .001$.

TABLE 3

GROUPS MEANS (LOGS) ON CONDITIONS

	I	II	III
Control	2.11	2.15	2.24
Experimental	2.22	2.31	2.29

to the transformed data and the resultant chi square of 10.43 was not significant at the .05 level.

Table 2 presents the results of the complete analysis of variance. For the central hypothesis only the Conditions × Groups interaction is relevant. The effects of Cards and Order of Presentation of the Cards are only of secondary interest.

The Groups × Conditions interaction mean square was .02736, giving an *F* of 3.44 with 2 and 60 *df*, significant at the .05 level. Thus, the subliminal stimulus conditions did result in a difference betweeen groups. The nature of this difference can be seen in Table 3 which gives the means of the two groups (in logarithm scores) for the three conditions. The control group showed a steady rise in productivity going from Conditions I to III. The experimental group showed a slightly greater rise in productivity from Condition I (blank) to Condition II (WRITE MORE) and showed a drop instead of a rise from Condition II (WRITE MORE) to Condition III (DON'T WRITE).

In order to examine more closely the effect of the subliminal stimulus, individual comparisons were carried out on the difference in mean change from one condition to another in the two groups.[2]

Three *t* tests were performed between the three Conditions. Only the difference between groups from Condition II to Condition III was significant ($t = 2.61$, $p = .025$ for a one-tailed test). Apparently the subliminal suggestion "DON'T WRITE" played the greater role in the interaction between groups and conditions. On Condition III, 13 of the 18 *S*s in the experimental group decreased their productivity (in accord with the subliminal suggestion) while only 4 of the 18 *S*s in the control group decreased their productivity.

Condition III consisted of 10 trials. On each successive trial the picture stimulus was exposed for a longer interval. One might expect the subliminal suggestion to have some kind of increasing effect as trials progress. Figure 1 contains plots of the mean and median (logarithms) numbers of words written on each trial for the two groups. There was considerable variability within groups on single trials causing the means to be somewhat distorted by extreme results. In terms of mean scores, it can be seen that the experimental group started on Trial 1 at a considerably higher level of productivity. This difference is probably a carry-over from Condition II. The large initial difference between the groups on Trial 1 disappears by Trial 3 as a result of the rapid increase in the Control group and the decrease in the Experimental group. Beyond this trial the groups do not markedly diverge. The medians present a somewhat different picture. The experimental group remains more productive up to the seventh trial, and from the seventh through the tenth trial drops below the control group. With the exception of an "end-spurt" in both groups, there is a tendency for the experimental group to *decrease* while the control group *increases* over the successive

---

[2] For instance, for the change from Condition II to III we have:

$$t = \frac{(\bar{Y}_{eII} - \bar{Y}_{eIII}) - (\bar{Y}_{cII} - \bar{Y}_{cIII})}{\sqrt{MS_{e}4/18}} \text{ with 60 } df$$

where $\bar{Y}_{eII}$ is the mean for the experimental group on Condition II, etc. These *t*'s were judged for significance using Tukey's Studentized Range test for multiple comparisons (Scheffe, 1953), the .05 level in this case being 2.40.

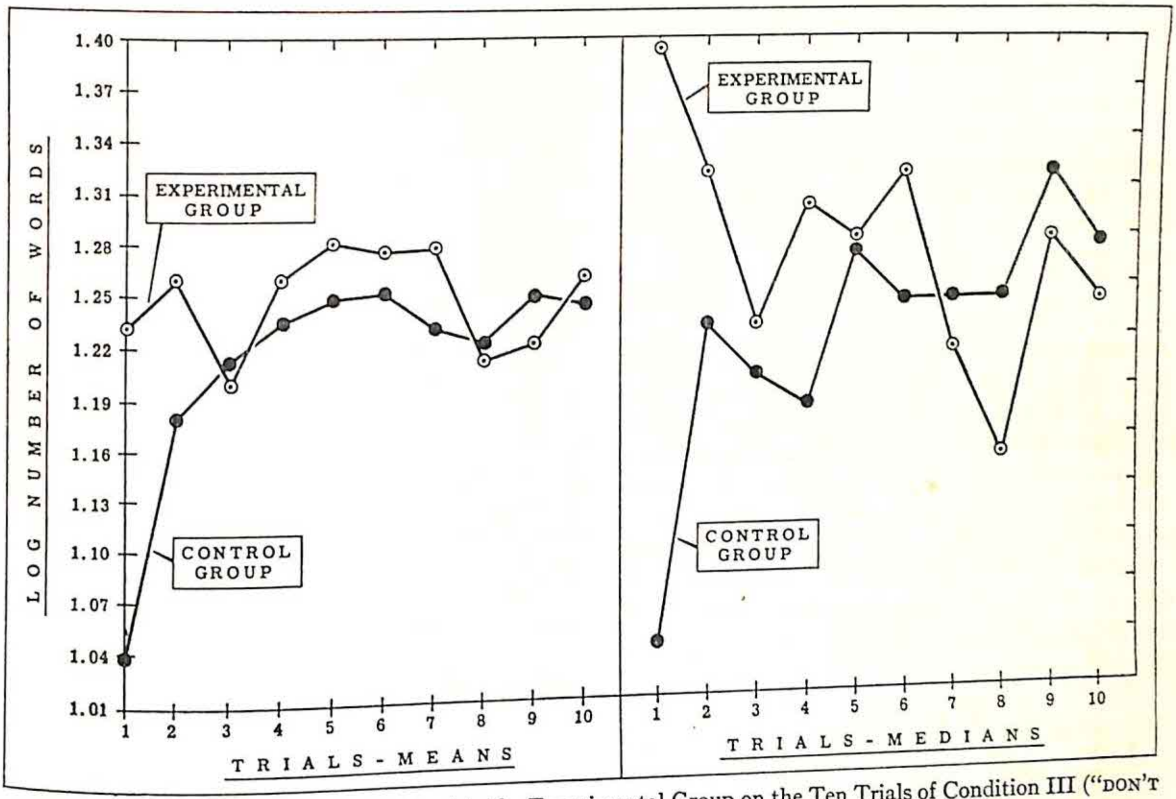


FIG. 1. Means and Medians (logarithms) in the Experimental Group on the Ten Trials of Condition III ("DON'T WRITE").

trials. Thus, there is some indication that the subliminal suggestion "DON'T WRITE" has a cumulative effect, when presented over a series of trials.

In order to examine the relationship between the *S*'s threshold and her response to the subliminal suggestion, a rank order correlation was performed between the percentage drop in productivity on Condition III in the Experimental group and the *S*'s threshold for the words "DON'T WRITE" obtained against a blank background at the conclusion of the experiment. The 18 experimental group *S*s were ranked from the greatest percentage increase to the greatest percentage decrease in productivity on Condition III, and thresholds were ranked from the highest to the lowest. The rho obtained between these two rankings was −.32. Although this value falls short of significance, the direction of the correlation suggests that high thresholds tended to be associated with low, or weak, response to the subliminal suggestion.

The results up to this point indicated that one of the subliminal suggestions had a significant influence on the productivity of the *S*s. The direction of the correlation between the threshold and the magnitude of this influence suggested that, perhaps, the influence of the suggestion increases as the *S* becomes more aware of it. A third group of *S*s was therefore tested, this time increasing the duration of the suggestion stimuli so that they would become supraliminal.

EXPERIMENT II

Method

Eighteen more student nurse *S*s composed the third group of the experiment. Their mean age was 19.94 ($SD = 1.30$). The procedure was the same as that of the previous experimental group except that this time the duration of the suggestion stimulus was increased to .5 sec., its onset beginning .02 sec. before the shut-off of the picture stimulus. It was assumed that this half-second duration would make the suggestion stimulus supraliminal for all *S*s. The threshold for the Condition III suggestion stimulus presented against a blank background had not exceeded .2 sec. for any of the *S*s in Experiment I.

Results

To the surprise of the experimenter, 10 of the 18 *S*s still did not report seeing either of the suggestion stimuli when they were questioned after the experiment. The remaining eight *S*s reported seeing one or both of the suggestion stimuli, during the experiment.

The *S*s who did not report seeing the suggestion stimuli were shown the succession of picture and suggestion stimuli with the instructions:

> Now watch carefully. There are some words following the picture. See if you can read them.

After being "set" all except one of the 10 *S*s were able to report the suggestion stimulus correctly. They were questioned as to why they thought they had not seen the stimulus previously. From their comments it was apparent that most of them were close to threshold but were not attending to the second stimulus. Four of the 10 were aware that words were being presented but could not read them. Their comments were as follows:

> A. I noticed words—saw "Don't."
>
> B. I couldn't read the words.
>
> C. I remember I saw little squares. . . blur in it. I thought it was the tag on the bottom of the picture. If I'd been paying attention I could have seen it.
>
> D. It looked like it might be the title of the picture. . . . Printing in the center of the card I was looking at.

Three *S*s were aware of only the flash. Their comments were:

> E. I noticed a flash. I guess it was because I was concentrating so hard on the picture. . . . I was annoyed at the light. . . .Just trying to get it out of the way more or less.
>
> F. I saw a flash but it was too quick.
>
> G. Flash . . . but I was just expecting pictures.

The remaining three *S*s did not notice anything but the pictures.

> H. I didn't expect to see any words.
>
> I. Was that there all the time?
>
> J. It wasn't that slow before.

Results for these 10 *S*s were analyzed as a second subliminal group on the assumption that they were closer to threshold (or even partially aware of the stimulus) than the .02 sec. subliminal group; the remaining eight *S*s were regarded as a "supraliminal" group. Comparing these two groups with the two groups from the previous experiment might enable one to determine the relationship between closeness to threshold, or degree of awareness, and effect of suggestion. It was assumed that the four groups lie on a continuum of awareness in the following order: Control group, Subliminal .02 group, Subliminal .5 group, Supraliminal group.

The variability introduced by the last two groups made a parametric analysis inappropriate even for log converted data (Bartlett's test yielded a chi square of 75.95). The statistical method used was the Kruskal Wallis One-Way Analysis of Variance by Ranks (Siegel, 1956). The percentage increases and decreases of the 54 *S*s in all four groups were ranked from the greatest decrease to the greatest increase. The mean ranks in each of the four groups on the percentage change from Conditions I to II and the percentage change from Conditions II to III can be seen in Table 4. Large percentage drops in productivity lower these means while large percentage increases raise them. The median percentage drop or increase in productivity is also indicated in this table.

The value of H, after correction for ties, for the change from Condition I to II was 2.61, which was not statistically significant ($df = 3$). The value of H for the change from Condition II to III was 13.33, which was significant below the .01 level. Since the mean ranks of the four groups were significantly different on the percentage change from Conditions II to III individual tests of significance were performed on differences among the mean ranks.[3] The z values resulting from these tests can be seen in Table 5. Both the .02 and .5 sec. Subliminal groups show a significantly greater drop in productivity on Condition III than the Con-

---

[3] For the difference between the mean ranks of any two groups, a and b,

$$z = \frac{\bar{R}_a - \bar{R}_b}{\sqrt{\sigma R^2\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}}$$

where

$$\sigma R^2 = \frac{N(N+1)(N-1) - \Sigma T}{12(N-1)}.$$

In this application, $N = 54$, and ΣT was 126. (It can be shown that if there were only two groups, $z^2 = H$). To test these zs for significance, the Tukey Studentized Range Test was again used, this time with $df = \infty$.

TABLE 4

MEDIAN PERCENTAGE CHANGES AND MEAN RANK CHANGES IN FOUR GROUPS

Group	N	Condition I to Condition II (WRITE MORE)		Condition II to Condition III (DON'T WRITE)	
		Med. Change (%)	M Rank	Med. Change (%)	M Rank
Control	18	+7	22.28	+15	37.42
.02 Subliminal	18	+20	28.89	−8	22.94
.5 Subliminal	10	+40	34.80	−12	16.35
Supraliminal	8	+21	27.00	+17	29.13
H			2.61		13.33
p			*NS*		<.01

TABLE 5

TESTS OF SIGNIFICANCE (z) BETWEEN MEAN RANKS OF FOUR GROUPS

Group	.02 Subliminal	.5 Subliminal	Supraliminal
Control	2.76*	3.40**	1.24
.02 Subliminal		1.06	.93
.5 Subliminal			1.71

* Significant below the .05 level.

** Significant below the .01 level.

trol group, but the difference between the Supraliminal group and the Control group is not significant. Although the .5 Subliminal group drops more in productivity than the .02 Subliminal group, the difference between these groups is not significant.

It would thus appear that the suggestion effect only appeared when the suggestion stimulus was subliminal. The .5 Subliminal group merely furnishes a replication of the results in the .02 subliminal group. Why was the supraliminal suggestion ineffective? A closer look at the Supraliminal group indicates that the suggestion did affect at least one of the *S*s in this group. Only three of the eight *S*s decreased their productivity on Condition III, but one of these *S*s complied in full with the suggestion by not writing at all during this condition! Her decrease was therefore 100% from the previous trial. A second *S* decreased 56%. These two *S*s showed more response to the suggestion than any of the other *S*s in the experiment. However, their response is masked in the group data by four other *S*s who showed large increases in productivity, or negative responses to suggestion. The *S* who fully complied with the suggestion said: "I know it said 'don't write,' so I tried to concentrate on the following ones, in case it said 'write.' I couldn't see any reason for the words being there unless they were directed for me."

The *S* who decreased by 56% felt that the words were there to influence her because she had read about "subliminal advertising" but she wasn't aware that her productivity had been affected by the suggestion.

The *S*s who were contrasuggestive, i.e., increasing instead of decreasing their productivity, found some interpretation for the stimulus which explained their disregarding it.

One *S* said: "I didn't know whether it meant stop writing or not so I just went on writing what I saw."

Another *S* said: "I was just surprised to see it there. I didn't notice it before. I was trying to look at the pictures and didn't pay too much attention to it. I thought it meant don't write at that time, not to write until the light went off."

Another *S* said: "My first impulse was that it was saying 'Don't write,' but due to your previous instructions, I went ahead. Then I thought it had to do with the picture; the man leaving the woman and saying 'Don't write.' "

A final question which the experimenter tried to answer was the question of whether the individual differences in response to the suggestion stimulus were related to personality differences in *S*s. A previous study by Zuckerman and Grosz (1958) demonstrated some relationship between two of the needs measured by the Edwards Personal Preference Schedule (EPPS) (Edwards, 1954) and response to the sway test (a measure of suggestibility). The EPPS was administered to eight *S*s from the two subliminal groups who had responded in a contrasuggestive direction on Condition III (i.e., increasing rather than decreasing their productivity), and to the eight *S*s from these groups who had shown the largest positive response to the suggestion (i.e., the largest percentage decreases in productivity). Comparison of these two groups on the 15 Murray needs measured by the test yielded only one significant difference. The contrasuggestive group was significantly higher on need Affiliation (p less than .01). Since one difference out of 15 comparisons may be due to chance, it may be concluded that no personality differences between these groups were revealed by the instrument used.

DISCUSSION

The data indicate that one of the two subliminal suggestions influenced the productivity of the *S*s in describing the pictures. It is not clear why the suggestion "Don't Write" was more effective than the suggestion, "Write More." The suggestion Don't Write seemed to have the effect of counteracting a practice effect, for the control group steadily increased their productivity going from Condition I to Condition III. One can speculate that there were two opposing motivations in the last condition: (*a*) the usual positive motivations of experimental *S*s for personal achievement or to please the experimenter; (*b*) a fatigue effect toward the end of the experiment resulting in a disposition to exert less effort. The suggestion Don't Write may have reinforced the increasing disposition to slow down. More work is needed to explore the relationship between subliminal suggestion and the intensities and types of needs or motives being reinforced.

The degree of "awareness" or recognition plays a crucial role in disputes over the phenomena of subliminal perception. Goldiamond (1958) feels that the phenomena can be explained in terms of "partial awareness" which the *S* cannot communicate because of the lack of quantitative specificity in the verbal indicators used to define "perception." This question can be treated in empirical terms by asking "what is the relationship between the subliminal suggestion effect and the 'degree of awareness' of the subliminal stimulus?" This question has been investigated in this experiment by: (*a*) examining the relationship between the magnitude of response to suggestion and the threshold for the suggestion stimulus and (*b*) comparing two subliminal groups (differing markedly in the time of exposure of the subliminal stimulus), a supraliminal group and a control group. There is some tendency for a greater suggestion effect in the *S*s operating below, but closer to threshold. However, neither the correlation between threshold and magnitude of response, nor the difference between the .02 sec. and .5 sec. Subliminal groups reached significance. Both of the subliminal groups did differ significantly from the control group, whereas the supraliminal group did not. *S*s who became completely aware of the suggestion seemed to respond in a dichotomous fashion: either conforming strongly with the suggestion or responding strongly in a contrasuggestive direction. The latter tendency was actually more prevalent. This points up the difference between the present experiment and those in which the *S* is given a discrimination task. When the *S* is looking for an answer to a problem, he will, of course, use the answer that the experimenter gives him, and the greater the intensity of the cue the more likely he will respond. However, when the subliminal cue is not directly relevant to the primary task on which the *S* is focusing, he is quite likely to reject or ignore it as an annoying distraction. This is what many of the supraliminal *S*s in this experiment apparently did. The "distracting" or "irrelevant" nature of the subliminal cue may account for the difficulty which the majority of *S*s had in seeing it even though it was exposed for a relatively long interval following the picture to which they were attending.

SUMMARY

Two experiments were performed to study the effect of subliminal and supraliminal suggestion on productivity, or number of words written in describing a picture. An experimental group was given a task of describing in writing a TAT picture presented tachistoscopically 10 times at increasing exposure levels. In the second condition a subliminal stimulus, the words WRITE MORE overlapped with the TAT picture for .02 seconds. In the third condition the subliminal suggestion was DON'T WRITE. A different TAT card was used for each condition and the cards were rotated among conditions in three orders so that each card was presented in each condition an equal number of times. The control group performed the same task on the three cards but instead of a subliminal stimulus only a blank flash occurred in Conditions II and III. There were 18 *S*s each in the experimental group and control groups. Significant differences between groups as a function of conditions were found. Whereas the control group increased their

writing rate steadily from Condition I to Condition III, the experimental group decreased their rate on Condition III.

A second experiment was performed to see the effect on productivity of the same suggestions at supraliminal levels. The exposure intervals of the suggestions were increased to .5 sec. Ten of the group of 18 tested under these conditions still could not report awareness of the suggestions, although some of them had "partial awareness." These *S*s showed the same effect as *S*s in the .02 sec. condition. The eight *S*s who recognized at least one of the suggestions did not show a consistent response to them. Some responded in the suggested manner, i.e., markedly increasing or decreasing their rate on request, while others ignored the suggestion or did not interpret it as "instructions." The net effect was closer to the response of the control group than to that in either of the subliminal groups. An hypothesis formed from these data is that subliminal suggestion may produce some effect in the region just below threshold. When the suggestion becomes supraliminal its distracting effect causes a contrasuggestive response in some *S*s. No marked personality differences (as measured by the EPPS) were found between suggestive and contrasuggestive *S*s in the two subliminal groups.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

EDWARDS, A. L. *Edwards Personal Preference Schedule. Manual.* New York: Psychological Corp., 1954.

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1958. Pp. 169–227.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

MCCONNELL, J. V., CUTTER, R. L., & MCNEIL, E. B. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, **13**, 229–242.

SCHEFFE, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

ZUCKERMAN, M., & GROSZ, H. J. Suggestibility and dependency. *J. consult. Psychol.*, 1958, **22**, 328.

(Received April 28, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

CONFUSION OF FIGURATIVE AND LITERAL USAGES OF WORDS BY SCHIZOPHRENICS AND BRAIN DAMAGED PATIENTS[1]

LOREN J. CHAPMAN[2]

University of Chicago

MANY clinical writers have commented on an aspect of the thinking disorder of schizophrenic and brain damaged patients which they describe as a confusion of literal usages of words with more figurative or metaphoric usages. The present study is an attempt to investigate such phenomena by means of objective measures.

Clinical writers have seldom defined the terms literal, figurative, and metaphoric, but have instead relied on the use of example to indicate their meaning. However, linguists have offered explicit definitions which should be of use to clinicians who are concerned with this area.

A figure of speech is usually defined by linguists as an extension of the applicability of a word. Figures of speech are often divided into such categories as metaphor, metonomy, and synecdoche (Sturtevant, 1947, pp. 137–139). Metaphor is usually defined as the extension of the applicability of a word in order to suggest a similarity or analogy between two objects, ideas, or events, e.g., "pull strings" for "use influence." Such metaphoric extensions may be: (*a*) from one sense modality to another, e.g., *loud* color; (*b*) from one object to another within the same sense modality, e.g., the use of the term *moon* to describe man-made satellites; (*c*) from a sense modality to a notion that ordinarily is not susceptible to direct sensory apprehension, e.g., *warm* feelings. Freshly coined and original metaphors, which include those of poetry and other similar literature, are readily recognized as such and usually require effort on the part of the listener to grasp their meaning. With continued usage, a metaphor becomes incorporated into the language to the point that it is seldom recognized as a metaphor, e.g., *high* pitch.

The present investigation is concerned with metaphors which are fairly well incorporated into everyday usage but are still recognizable as metaphors. It is primarily concerned with the third kind of shift listed here, shifts from a sense modality to the description of subjective judgments, emotional states, or the quality of interpersonal interactions.

Clinical writers have described the thinking disorder in terms of a disturbance in the use of this type of metaphor both in schizophrenics (Arieti, 1955; Bateson, 1956; Bleuler, 1950; Goldstein & Scheerer, 1944; Muncie, 1937; Storch, 1924) and organics (Goldstein, 1948; Head, 1926; Jakobson & Halle, 1956). All of these writers attribute both inappropriately literal and inappropriately metaphoric utterances to patients. However, most of these writers (except Bateson and Muncie) state that the apparent metaphoric utterances are seldom or never truly metaphoric, but are, instead, meant in a literal way by the patient. For example, Storch (1924, pp. 5–9) construed a schizophrenic's reference to a person's train of thought as a "thread" and his will as a "spark" as referring to sensory images of a physical spark and a physical thread without any metaphoric intent on the patient's part.

Such clinical views seem to imply that if the patients were presented with words in context and asked to indicate their meaning, they would often misinterpret figurative usages as if they were literal but would not commit the reverse error of misinterpreting literal usages as if they were figurative. That is, such a patient might make a literal misinterpretation of the statement, "The only way Tom can get appointed to West Point is to pull strings," as meaning, "It is necessary for him to grab a cord in each hand and yank it"; however, he would not commit the converse *figurative* misinterpretation of "pull strings" in the statement, "The

---

[1] This study was supported by a research grant (M-1416) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

The author is indebted to Alvin Burstein for his valuable suggestions in the preparation of the manuscript, to Dorothy Day for assistance in the collection and analysis of the data, and to Catherine Becker for assistance in analysis of the data.

[2] Now at the University of Kentucky.

only way for Jim to lace his boots is to pull strings." Nevertheless, Bateson (1956) and Muncie (1937) specifically state that schizophrenics commit this latter kind of error. Muncie believes that it is more prominent in schizophrenia than excessive literalness but that both kinds of misinterpretation occur. For Bateson, the tendency of schizophrenics to make both kinds of misinterpretation is a central feature of his theory of schizophrenia.

In the present investigation, schizophrenics, brain damaged patients, and normal *S*s were compared on these two kinds of misinterpretation.

PROCEDURE

Instrument

A multiple-choice paper-and-pencil test was constructed. The following examples are representative of the items used in the test:

9. David turned yellow when he faced the enemy.
   This means:
   A. David became cowardly.
   B. David became hungry.
   C. David's skin became discolored.
31. Miss Bailey's illness turned her yellow.
   This means:
   A. Miss Bailey became cowardly.
   B. Miss Bailey became hungry.
   C. Miss Bailey's skin became discolored.

In Item 9 above, the correct answer is, of course, "David became cowardly." Choosing "David's skin became discolored" would reflect an inappropriately literal interpretation of the word "yellow." In Item 31, the reverse is true; "Miss Bailey's skin became discolored" is the correct answer, and "Miss Bailey became cowardly" is inappropriately figurative.

The third alternative, which interprets yellow to mean hungry, is neither correct nor a predicted error due to misinterpretation. But inclusion of this alternative has a special use in the design. Errors are made for many reasons besides the kind of misinterpretation under investigation, and the bases of these errors are mostly unidentified. For example, many schizophrenics are noncooperative and respond carelessly, mark randomly, mark by position, or mark in response to other characteristics of the test which are difficult for the examiner to identify. If such unidentified errors are equally distributed among the three alternatives, then a score of predicted error markings minus "unidentified" error markings should be a relatively pure measure of the predicted error tendency.

Another way of putting it is that there should be, by chance, as many errors made for unidentified reasons hidden among the predicted misinterpretation markings as there are markings of these "unidentified" error alternatives themselves. Therefore, we may obtain for each *S* a relatively pure measure of literal misinterpretation if we subtract from the number of his literal error markings the number of his "unidentified" error markings. This will be called the "corrected literal score." The "corrected figurative score" is obtained by following the same principle.

The instrument was a 40-item paper-and-pencil test[3] which consisted of 20 items which permit literal misinterpretations and 20 items which permit figurative misinterpretations. The items were constructed in pairs, as in the example above, with one item testing for a literal misinterpretation of a given word or phrase and the other for a figurative misinterpretation of that same word or phrase. In each case, the "correct" interpretation was determined by the context. The 20 figurative and the 20 literal items were intermixed in random order, with the restriction that the two items testing the literal and figurative aspects of the same phrase were separated by several other items. The order of the three kinds of alternatives was randomized among items with the restriction that the two items within each pair had the same order.

Subjects

Three groups of *S*s were tested[4]: 36 normal *S*s, 41 brain damaged *S*s, and 51 schizophrenics. The normal *S*s consisted of 23 medical and surgical patients at the West Side VA Hospital, Chicago, Illinois, and 13 Chicago firemen. No normal *S* was used for whom there was any indication either of psychiatric or neurological disorder.

The schizophrenic group, patients at Chicago State Hospital, had all received a hospital diagnosis of chronic schizophrenia, with their first hospitalization for psychosis occurring at least five years previously and with continuous hospitalization for at least three months previous to testing. Almost all of the schizophrenic patients had spent most of the last 10 years in the hospital. Other prerequisites for selection were absence of any indication of either cerebral organic involvement or mental deficiency, no current drug therapy, and no electric shock therapy within the past three months. No selection was attempted by diagnostic subtype.

The brain damaged patients consisted of 29 patients at Hines VA Hospital and 12 patients at Chicago State Hospital. The group included 13 diagnoses of cerebral vascular disorder, 6 of alcoholism, 6 of multiple sclerosis with cerebral disorder, 3 of brain tumor, 2 of head injury, 1 of Huntington's Chorea, 1 of syphilitic paresis, and 9 of unspecified cerebral damage. The organics, as a group, had been hospitalized for a shorter term than the schizophrenics. About one-third of the organics had entered during the previous three months, and none had been hospitalized longer than five years.

---

[3] Copies of the test may be obtained from the author, or may be ordered from the American Documentation Institute. Order Document No. 6237, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[4] The author is indebted to Kalman Gyarfus and J. Herbert Maltz of the Chicago State Hospital for permission to test schizophrenic patients for this experiment, to Gabrielle P. Ferrazzano for permission to test patients at the USPHS Hospital, to Roy Brener for assistance in obtaining patients at Hines Hospital, to Meyer Williams for assistance in obtaining *S*s at the West Side VA Hospital, and to Robert J. Quinn for his help in obtaining the cooperation of Chicago firemen.

Since these patients were being tested in order to study the nature of cognitive disturbance when it results from brain injury, only those patients were included who appeared by clinical interview to have some intellectual loss or disturbance. This decision was usually made on the basis of the patient's difficulty in communication, serious disturbance of memory, or severe lack of alertness as observed either in a clinical interview or in his performance on the CVS, a brief intelligence scale (Hunt, French, Klebanoff, Mensh, & Williams, 1948). No patient was included in the brain damaged group who appeared also to be schizophrenic.

The mean age was 43.4 years for the schizophrenics, 44.7 for the brain damaged, and 43.1 for the normals, while mean years of schooling was 10.2 for the schizophrenics, 10.2 for the brain damaged, and 10.6 for the normals. Although the groups did not differ significantly with respect to age or education, the schizophrenic group was lower ($p < .001$) than the organics on functioning IQ as determined by the CVS , and the organic group was lower ($p < .01$) than the normal group, presumably due to the disruptive effects of the schizophrenic and organic pathologies. The mean IQ was 86 for the schizophrenics, 104 for the organics, and 113 for the normals.

The *S*s were assumed to be of approximately equal premorbid intelligence because of the equivalent education. An additional approximation of premorbid intelligence was obtained from the item number of the last vocabulary word answered correctly. As suggested by Becker (1956), this appears to be a better estimate than the total number of vocabulary words answered correctly because one effect of psychopathology is the production of uneveness in performance. On this measure, the schizophrenics scored a mean of 23.3, the organics 26.2, and the normals 27.7. None of the differences was significant.

Results

All between-group differences were evaluated by a two-tailed Mann-Whitney test and all within-group differences by a two-tailed Wilcoxon test.

Table 1 lists the median error scores for each of the three groups. Each of the three groups marked significantly more literal misinterpretations than "unidentified" error alternatives and significantly more figurative misinterpretations than "unidentified" error alternatives ($p < .001$ in each case).

TABLE 1

Median Scores of Predicted and Unidentified Errors

	Schizophrenic	Brain damaged	Normal
Figurative error items			
Figurative errors	2.6	1.9	1.0
Unidentified errors	.3	.2	.0
Corrected Figurative error score (Figurative minus Unidentified)	1.4	1.4	.9
Literal error items			
Literal errors	5.3	.6	.4
Unidentified errors	.7	.1	.0
Corrected Literal error score (Literal minus Unidentified)	3.1	.4	.4

Comparison of schizophrenic and normal performance. By corrected literal score the schizophrenics were more susceptible to literal misinterpretations than the normal *S*s ($p < .001$). However, there was only a trend towards a difference on the corrected figurative score ($p < .08$). It would appear, therefore, that the tendency toward literal misinterpretations is greater than that toward figurative misinterpretations by schizophrenics on this test. By comparing, for the schizophrenic group alone, the corrected literal score with the corresponding corrected figurative score, it was found that the literal misinterpretations exceeded the figurative ($p < .01$). The normal *S*s showed a trend toward the opposite error preference, but not significantly so ($p = .20$).

These findings tend to support Bateson's statements, in that both kinds of misinterpretations tend to be made more often by schizophrenics than by normal *S*s. The fact that the normal *S*s also chose the misinterpretations more often than the unidentified errors shows that the misinterpretations are not unique to pathological patients, but rather that their greater frequency in schizophrenics is an exacerbation of a normal error tendency. The greater prevalence of the literal misinterpretations is in keeping with the observations of Arieti (1955), Bleuler (1950), Goldstein (1944), and Storch (1924). This difference in frequency between two kinds of misinterpretation on this test for a single group of *S*s, however, may be a function of the particular literal and figurative items offered. Therefore, any finding concerning comparative susceptibilities of any one group to the two kinds of misinterpretations on this test may be lacking in generality. Nevertheless, any finding of a difference *between* groups on this test is not subject to this stricture because the same items were given to all *S*s.

Comparison of brain damaged patients and normal Ss. Unlike these schizophrenics, the brain damaged patients showed no tendency toward more literal misinterpretations, as measured by the corrected error score, than the normal *S*s ($p = .89$). They showed only a trend toward a greater number of figurative misinterpretations than normals ($p = .11$).

Comparison of the brain damaged patients' preference for one or the other kind of misinterpretation, as measured by the corrected scores, showed that they more often make figurative misinterpretations than literal misinterpretations ($p < .05$). The normal *S*s, it will be recalled, showed no significant preference, while the schizophrenics showed the opposite preference.

Comparison of schizophrenic and organic brain damaged patients. The schizophrenics made more total errors than the brain damaged patients on both tests, but this may merely reflect differing degrees of pathology between the particular samples of patients tested. Comparisons of gross error tendency, which are so commonly reported in the literature, do not appear to be meaningful on any single test unless the two groups are equated in some way on over-all pathology—an extremely difficult task.

The schizophrenics and the organics can be meaningfully compared, however, on their relative preferences for the two kinds of misinterpretations. The schizophrenics, as mentioned above, made more literal misinterpretations than figurative, while the brain damaged patients showed the opposite preference. This difference was more directly examined by computing for each *S* the difference between his two corrected error scores. On this measure, the schizophrenics and organics differed ($p < .001$), and they were on opposite sides of normal *S*s. The schizophrenics differed significantly from the normals on this score, ($p < .01$), but the brain damaged patients did not ($p = .26$).

Although the schizophrenics and brain damaged patients differed in the kind of misinterpretation preferred, some of the patients in each group showed a preference opposite to that of their group as a whole. For the schizophrenics, 61% preferred the literal misinterpretation, 25% preferred the figurative, and 14% showed no preference. For the brain damaged patients, the preferences were 29% for the literal misinterpretations, 54% for the figurative, and 17% with no preference. It is interesting to conjecture whether the *S*s with atypical scores in each group may differ in other ways from the main body of patients. However, these atypical scores may have arisen merely from the unreliability of the test. For the schizophrenic patients, the Kuder-Richardson (Formula 20) reliabilities were .47 for the corrected figurative error score and .55 for the corrected literal error score. The corresponding values for the brain damaged patients were .49 and .43, respectively.

SUMMARY

A test was designed to measure the tendency to make inappropriately literal and inappropriately figurative misinterpretations of written statements. The schizophrenic patients made significantly more literal misinterpretations than figurative, while brain damaged patients showed the opposite effect.

A score of literal misinterpretation minus figurative misinterpretation distinguished the two groups and placed them on opposite sides of the normal *S*s.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.

BATESON, G., JACKSON, D. N., HALEY, J., & WEAKLAND, J. Toward a theory of schizophrenia. *Behav. Sci.*, 1956, **1**, 251–264.

BECKER, W. C. A genetic approach to the interpretation and evaluation of the process reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: Int. Univer. Press, 1950.

GOLDSTEIN, K. *Language and language disturbances.* New York: Grune & Stratton, 1948.

GOLDSTEIN, K., Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944.

HEAD, H. *Aphasia and kindred disorders of speech.* Vol. 1. New York: MacMillan, 1926.

HUNT, W. A., FRENCH, ELIZABETH G., KLEBANOFF, S. G., MENSH, I. N., & WILLIAMS, M. Clinical possibilities of an abbreviated individual intelligence scale. *J. consult. Psychol.*, 1948, **12,** 171–173.

JAKOBSON, R., & HALLE, M. *Fundamentals of language.* The Hague: Mouton, 1956.

MUNCIE, W. The psychopathology of metaphor. *Arch. Neur. Psychiat., Chicago*, 1937, **37,** 796–804.

STORCH, A. The primitive archaic forms of inner experiences and thought in schizophrenia. *Nerv. ment. Dis. Mongr. Ser.*, 1924, No. 36.

STURTEVANT, E. H. *An introduction to linguistic science.* New Haven: Yale Univer. Press, 1947.

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

LEARNING A SOCIAL STRUCTURE[1]

CLINTON B. DE SOTO
Johns Hopkins University

For some years social psychologists have shown increasing interest in mathematical properties of social structures, ranging from investigations of the extent to which sociometric choices are mutual to experimental manipulations of the centrality of communication networks. This interest has been accompanied, however, by a curious lack of concern with the subjects' (*S*s') views of these properties. For example, in the studies of mutuality of sociometric choices, the *S*s are not ordinarily asked if *they* expect the choices to be mutual. Only recently, in such work as Tagiuri's (1958) generalization of sociometry, relational analysis, and Heider's (1958) discussion of naive psychology, has there been serious concern with the *S*s' perception of such properties.

Apparently psychologists have assumed that *S*s cannot or do not appreciate these properties. Yet folklore is full of essentially mathematical statements about interpersonal relations: "The way to get someone to like you is to like him," "You can't serve two masters," "There's always a faster gun somewhere." And recent questionnaire studies (De Soto & Kuethe, 1958, 1959) have demonstrated that *S*s ascribe such properties rather freely to interpersonal relations, sometimes accurately, sometimes inaccurately.

Historically social psychologists have repeatedly been compelled to recognize the importance of cognitive processes in man, to the need, for example, for crediting people with role expectations as well as role behaviors. This fact makes it appear unwise to ignore their readiness to assign mathematical properties to interpersonal relations. It makes it important to look for effects of such attributions of properties in activities besides answering questionnaries. There are in the literature many findings which are suggestive of such effects. For example, Kogan and Tagiuri (1958) found that people expect those they choose to choose one another, thus evidencing, for a special case, their expectation that sociometric choices will be mutual.

One task in which *S*s' tendencies to assign properties to interpersonal relations might be expected to have especially prominent effects is that of learning a social structure. Learning a social structure means learning the relations among a set of people. It is a task that usually faces a person when he enters a social group, and it can be quite difficult as indexed by the information he must assimilate. It is a reasonable prediction that his assignments of properties to the relations which form the structure, operating as expectancies or hypotheses, will facilitate or hinder his learning of the structure according to their validity. The present study was designed to verify this general prediction and also to examine in more detail some of the psychological processes involved.

The study was basically a verbal learning experiment in which *S*s learned miniature social structures. The dependent variable was the number of trials required. There were two independent variables: the interpersonal relation that formed the structure, and the mathematical properties of the structure. In their questionnaire study of the properties assigned to various interpersonal relations, De Soto and Kuethe (1959) found that the relations fell into two main classes: those regarded as asymmetric, which were also regarded as transitive, and those regarded as symmetric, some of which were also regarded as transitive and some of which were not. For the present study, the relation "influences" was chosen to exemplify the relations people expect to be asymmetric and transitive, "likes" was chosen to exemplify those expected to be symmetric and transitive, and "confides in" was chosen to exemplify those expected to be symmetric but not transitive.

Structures characterized by eight different sets of mathematical properties were used. Four-man structures were used because they were the largest *S*s could learn with relative

---

[1] This work was supported in part by Grant NSF-G4827 from the National Science Foundation. The writer is grateful to Martin N. Narun and Daniel M Glasner for aid in the collection of data.

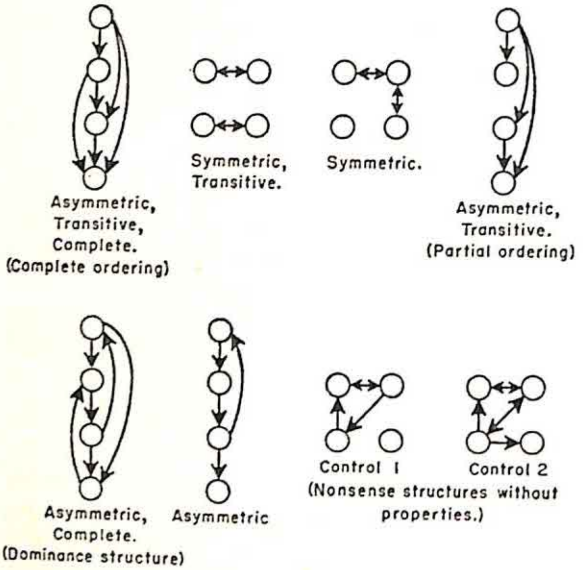


FIG. 1. Structures used in the learning experiment.

ease in pilot studies. The structures are depicted in Fig. 1, with circles representing people and arrows a relation. An arrow going from one circle to another means the first person stands in the given relation to the second; if the relation is "influences," it means the first person influences the second. The requirements for the various properties may be defined informally as follows: For symmetry, the arrows must go in both directions, or not at all, between a pair of individuals. For asymmetry, the arrows must never go in both directions between a pair of individuals. The requirement for transitivity is more complex. If there is an arrow from Individual *a* to Individual *b*, and another from Individual *b* to Individual *c*, then, for the transitive property to hold, there must be one from *a* to *c*. For the complete property to hold, there must be at least one arrow connecting the individuals in all possible pairs. More formal definitions of these and other properties are given by Coombs, Raiffa, and Thrall (1954) and Kershner and Wilcox (1950). The names given to some of the structures in Fig. 1 are based on their properties.

The task given each *S* in the experiment was to learn one of the structures. He did not see a drawing of the structure, however, nor was the relation represented by abstract arrows. Instead, the arrows were replaced by one of the flesh-and-blood relations, "influences," "likes," or "confides in," and he had to learn, by repeated trial and error, whether or not Jim likes Ray, whether or not Ray likes Stan, and so on for the 12 ordered pairs of names which exist in a four-man structure. This arrangement of the task avoided unintended suggestions to the *S* about properties of the structure and approximated the real-life case where people learn a social structure by observing the pair-wise interactions among its members rather than from devices like organizational charts.

An experimental condition in this study was a combination of one of the abstract structures and one of the interpersonal relations. There were 19 experimental conditions. Each of the first three structures in Fig. 1 was formed from each of the three relations, producing 9 conditions, and the remaining five structures were formed from the relations "influences" and "likes," producing 10 more conditions. The first 9 conditions were the basic experimental ones, since a crucial interaction was predicted for them. On the basis of the conclusion by De Soto and Kuethe (1959) that college students (used as *S*s in that study and this) are alert to power based orderings, it was predicted that they would expect the relation "influences" to be asymmetric and transitive and, possibly, complete and would therefore learn a complete ordering based on "influences" more readily than one of the symmetric structures, while for "likes" and "confides in," which *S*s expect to be symmetric, the converse would be true. The remaining conditions were included to provide further analytical possibilities.

METHOD

The experimental method resembled that of traditional paired-associates learning experiments. The *S*s, male undergraduates at Johns Hopkins University, were assigned randomly, nine to each of the 19 experimental conditions, and were run individually.

Their task is described well by the instructions given them:

> This experiment is on the learning of interpersonal relationships. There are four people involved, Bill, Jim, Ray, and Stan, and your job is to learn who likes who among these people.
>
> On the upper side of each of these cards are printed the names of two of the four people. In each case, the first person either does or does not like the second. If he does, it says so on the other side of the

card. If he doesn't, nothing is printed on the other side of the card.

When I tell you to begin, look at the names on the top card. If you think the first person likes the second, say so. If you don't think he does, say so. For example, if the first name were Sam and the second were Charlie, you would say either "Sam likes Charlie" or "Sam doesn't like Charlie." Then, turn the card over and see whether or not he does, and place the card on a new pile.

Go through the rest of the cards the same way, and then I will shuffle them and you can go through them again. As we go through the cards again and again, you should learn better and better who likes who.

Take your time. Don't rush.

The above are the instructions given to all *S*s who learned a social structure based on "likes." The instructions given to *S*s who learned a social structure based on "influences" or "confides in" were identical in form except that the appropriate one of the following definitions was added: "To influence someone means to have some degree of control over what he does"; "To confide in someone means to have complete faith and confidence in him so that you trust him with secrets or rely on him for help."

Each *S* worked with a deck of 12 cards appropriate for his experimental condition.

RESULTS

The difficulty of learning the various social structures was measured by the mean number of trials required by the *S*s to learn them to a criterion of two successive correct trials. These means are shown in Table 1. The standard errors of the means, highly correlated with their values, range from 1.06 to 4.27.

The hypothesized interaction between relations and structures is evident in Table 1. The *S*s learned a complete ordering based on "influences" very readily, but had somewhat more difficulty with symmetric structures based on "influences." They had difficulty learning a complete ordering based on "likes" or "confides in," but learned symmetric structures based on these relations easily. An analysis of variance performed on the data for these nine conditions yielded an interaction mean square significant at the .01 level.

TABLE 1

MEAN TRIALS TO CRITERION

Structure	Relation Influences	 Likes	 Confides in
Asymmetric, transitive, complete (complete ordering)	8.9	14.7	16.2
Symmetric, transitive	12.9	10.9	9.1
Symmetric	11.7	13.0	10.9
Asymmetric, transitive (partial ordering)	12.1	14.7	
Asymmetric, complete (dominance structure)	26.4	23.7	
Asymmetric	16.3	29.1	
Control 1	16.1	20.6	
Control 2	16.7	21.0	

Examination of the remaining ten means shows that the partial ordering was learned relatively easily, at least when based on "influences," but the other conditions were all difficult, some of them exceedingly so. Particularly striking is the finding that the asymmetric and complete structure based on "influences" took substantially longer to learn than any of the nonsense structures, rather than showing facilitation from the fact that its properties were partly as expected by the *S*s.

DISCUSSION

The experimental results clearly support the general hypothesis that people's assignments of mathematical properties to an interpersonal relation influence the ease with which they learn a social structure based on the relation, and they invite analysis of the psychological processes involved.

Perhaps the most remarkable finding is this: although the combination of the asymmetric, transitive, and complete properties (complete ordering) greatly facilitates the learning of a social structure based on "influences," the combination of only the asymmetric and complete properties (dominance structure) seems to make the learning task even more difficult than it is for structures without properties. The *S*s failed completely to profit from the latter combination of properties, although in principle it too could be a valuable aid to learning a structure. Why?

It appears an inescapable conclusion that for the *S*s these properties were not independent of one another, even though they are independent mathematically, and even though *S*s are able to ascribe them separately to relations when questioned (De Soto & Kuethe, 1958, 1959). De Soto and Kuethe noted a tendency for *S*s usually to ascribe the transitive property whenever they ascribed the asymmetric property to a relation. Evidently this tendency to couple the properties is so powerful that it prevents people from using the properties in isolation when learning a social structure. People expect mathematical properties in social structures, to be sure, but they expect them in bundles, and incomplete bundles are of no use to them.

It is as if the *S*s had a theory about the social structure, a mathematical model for it which was very helpful when its properties corresponded with those of the social structure, but which had to be applied in an all-or-nothing manner in the sense that it was not readily abridged in the face of some lack of correspondence and became instead a hindrance. Such a theory, marked by certain essential properties but doubtlessly skeletal and sketchy in other respects, seems most aptly called a *schema*, roughly following Bartlet (1932) and Woodworth (1938).

The present data, then, might be interpreted to indicate that the *S*s possessed an ordering schema which they automatically applied to a social structure based on the relation "influences," although not to one based on "likes" or "confides in." "Influences" might be said to tap or arouse people's ordering schema. This ordering schema, the residue perhaps of countless experiences with social and nonsocial orderings in the world, seems to be marked by the properties of asymmetry, transitivity, and, possibly, completeness. Thus the *S*s expected these properties in a social structure based on "influences," but not as independent expectations; rather, as derivatives of the overriding expectation that the structure was an ordering. Perhaps the dominance structure was more difficult than the nonsense structures because it perpetually titillated and tantalized the *S*s' ordering schema, preventing the suppression of it that occurred for the nonsense structures.

Evidence that more sophisticated students of social structures also possess and can be victims of an ordering schema is found in many writings about peck rights among birds. Birds establish peck rights on the basis of pair-wise encounters, a process which insures that the eventual social structure is asymmetric and complete, but not that it is transitive. The birds, being simple-minded creatures not possessed of an ordering schema, thus innocently build a social structure that puzzles and provokes more thoughtful creatures who write about the "pecking out of order," "curious exceptions," "irregularities," "imperfect orderings," "lack of a hierarchy," "polygonal dominance," "cyclic pecking," "triangular pecking," and "complex, triangular hierarchies" (Allee, 1931; Collias, 1944; Murchison, 1935; Murphy, Murphy, & Newcomb, 1937; Schjelderup-Ebbe, 1935). People consider this primitive and simple social structure surprising and complex because they expect an ordering. To turn the thing around, it may be that the expectation of an ordering tends to *produce* an ordering, that human social groups would show fewer orderings and more dominance structures and even property free structures if it were not for the shared expectation of orderings.

The finding that the partial ordering was learned quite readily when based on "influences," with the *S*s' ordering schema presumably aroused, suggests that the property of completeness is a less integral part of the ordering schema than the properties of asymmetry and transitivity. However, any partial ordering can be mathematically treated as made up of two or more complete orderings (Coombs & Kao, 1955; Kershner & Wilcox, 1950), and it is possible that the *S*s accomplished essentially this analysis of the partial ordering.

The property of symmetry seems to be fairly easy to detect and utilize even when it is not expected, as in the structures based on "influences." The results suggest that for "likes" and "confides in" a symmetric and transitive structure is learned easier than a merely symmetric one, but the differences are not large, and it should be noted that the so-called symmetric and transitive structure of Fig. 1 is only *vacuously* transitive (Kershner & Wilcox, 1950), having too few members to be truly transitive. Thus, it is impossible to say more than that a lack of transitivity is not the severe handicap in learning symmetric structures that it is in learning asymmetric ones. This conclusion is consistent with the view expressed earlier (De Soto & Kuethe, 1959) that the primary property ascribed to these relations is symmetry, and that transitivity is ascribed secondarily, only after long experience which teaches that relations like "likes" are commonly transitive, dividing people into cliques or friendship groups. The grouping schema, as applied to interpersonal relations at least, is a higher development of the symmetry schema.

SUMMARY

By means of a learning experiment it was shown that a social structure is easiest to learn when it possesses the mathematical properties people expect in it. People attribute the properties of asymmetry, transitivity, and completeness (the properties of an ordering) to the relation "influences." If a social structure formed of the relation "influences" has these properties, they learn it readily. If it is symmetric, or if it has no mathematical properties, they have more difficulty; and if it is asymmetric and complete but not transitive (a dominance structure), they have great difficulty. Similarly, structures based on "likes" or "confides in" are learned most readily when they are symmetric, as people expect them to be. The findings were interpreted as reflecting the operation of schemas in the *S*s.

REFERENCES

ALLEE, W. C. *Animal aggregations: A study in general sociology.* Chicago: Univer. Chicago Press, 1931.

BARTLETT, F. C. *Remembering.* Cambridge: Cambridge Univer. Press, 1932.

COLLIAS, N. E. Aggressive behavior among vertebrates. *Physiol. Zool.*, 1944, **17**, 83–123.

COOMBS, C. H., & KAO, R. C. Non-metric factor analysis. *Engng. res. Bull.* No. 38. Ann Arbor: Univer. Michigan Press, 1955.

COOMBS, C. H., RAIFFA, H., & THRALL, R. M. Some views on mathematical models and measurement theory. *Psychol. Rev.*, 1954, **61**, 132–144.

DESOTO, C. B., & KUETHE, J. L. Perception of mathematical properties of interpersonal relations. *Percept. mot. Skills*, 1958, **8**, 279–286.

DESOTO, C. B., & KUETHE, J. L. Subjective probabilities of interpersonal relationships. *J. abnorm. soc. Psychol.*, 1959, **59**, 290–294.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

KERSHNER, R. B., & WILCOX, L. R. *The anatomy of mathematics.* New York: Ronald, 1950.

KOGAN, N., & TAGIURI, R. Interpersonal preference and cognitive organization. *J. abnorm. soc. Psychol.*, 1958, **56**, 113–116.

MURCHISON, C. The experimental measurement of a social hierarchy in *gallus domesticus:* I. The direct identification and direct measurement of social reflex no. 1 and social reflex no. 2. *J. gen. Psychol.*, 1935, **12**, 3–39.

MURPHY, G., MURPHY, L. B., & NEWCOMB, T. M. *Experimental social psychology.* (rev. ed.) New York: Harper, 1937.

SCHJELDERUP-EBBE, T. Social life of birds. In C. Murchison (Ed.), *Handbook of social psychology.* Worcester, Mass.: Clark Univer. Press, 1935.

TAGIURI, R. Social preference and its perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Cal.: Stanford Univer. Press, 1958.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Early publication received October 22, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

REPEATED FRUSTRATIONS AND EXPECTATIONS IN HOSTILITY AROUSAL[1]

LEONARD BERKOWITZ

University of Wisconsin

IN THEIR now classic monograph, *Frustration and Aggression*, Dollard, Doob, Miller, Mowrer, and Sears (1939) hypothesized that the total strength of the "instigation to aggression" within an individual is a positive function of three aspects of the frustrations he had suffered: "1) the strength of the instigation to the frustrated response, 2) the degree of interference with the frustrated response, and 3) the number of frustrated response-sequences" (p. 28). Even accepting Miller's later admonition that these propositions refer to the *instigation* to aggression (for simplicity, we might say "aggressive drive") and not aggressive *acts* (1941), many critics still objected to their failure to differentiate among qualitatively different classes of frustrations. For example, Maslow (1941) maintained only threatening frustrations give rise to aggression, while mere deprivations presumably do not. In the same vein, Pastore (1952) proposed that arbitrary frustrations are more likely to increase aggressive drive than are nonarbitrary frustrations. One reason for this, he suggests, is that the arbitrary frustrations involve the nonfulfillment of an expectancy, and aggressive reactions are said to result primarily when the expected attainment of a goal is blocked. Bateson presented an essentially similar hypothesis in his discussion of cultural factors affecting the frustration-aggression relationship (1941). The Balinese apparently do not learn to expect strong satisfactions, supposedly because of the frequent frustrations imposed upon them by their mothers in the course of their socialization. Therefore, Bateson argued, "they are infinitely willing to suffer interruption" (p. 353).

Dollard and his colleagues generally did not consider the operation of expectancies in affecting the reactions to frustrations, and this lack is particularly felt in their analysis of frustrated response-sequences. There is no unequivocal support for their principle that the strength of the aggressive drive increases with the number of specific frustrations. The only evidence they give for the hypothesis is anecdotal, and experimental tests of the proposition have yielded inconsistent results. Otis and McCandless (1955) conducted an investigation with preschool children placed in an eight-trial frustration task. The *S*s showed significant increases in "aggressive-dominant" behavior scores from the first four to the last four frustration trials, and showed reliable decreases in a class of behavior found to be incompatible with aggression, "submissive-complaisant" behavior. In contrast to this support for the proposition, McClelland and Apicella (1945) obtained negative findings in an experiment employing college students on only two frustration trials.

These opposing results can be reconciled by raising the possibility of differences in the development of understanding or expectations following the first frustrations. The person who anticipates further frustrations after the first disappointment may not be as upset when they occur as one who does not expect them. Conceivably, then, to the extent that the Otis-McCandless results are correctly interpreted as showing an increase in overt hostility with repeated frustrations, their *S*s may not have expected these continued blockings of their goal attempts. The McClelland-Apicella *S*s, being older and more sophisticated, presumably did develop these anticipations.

Findings consistent with this reasoning were reported in an earlier paper by the present writer (Berkowitz, 1960a). The *S*s received two acts from their partners. The first of these was either friendly or unfriendly in tone and the second (in a factorial design), although of a qualitatively different nature, also was either friendly or hostile. The data showed that the partner's second act had its greatest effect on *S*'s attitude toward him when it was

---

[1] This research was supported by Research Grant M1540(C1) from the National Institute of Mental Health, United States Public Health Service. The invaluable assistance of Thomas Murphy, Woodrow Call, and Diane Diamond, who collected the present data, is gratefully acknowledged.

inconsistent with the earlier behavior. That is, if the second act was relatively aggressive in manner, it provoked the greatest increase in unfriendliness toward the partner (*P*) when it had been preceded by the friendly behavior. (However, this was but one case of a more general principle: the greatest change toward friendliness following the second friendly act occurred when this behavior came after hostility from *P*.) It may be that the *S*s were relatively indifferent to the aggressive (or friendly) second act when this behavior was consistent with the earlier behavior because such actions were more or less expected of *P*.

The first experiment in the present report employed a procedure virtually identical to that used in the earlier investigations, and essentially the same results should be obtained. Assuming that the initial act from *P* gave rise to an hypothesis about him, there should be a greater resulting change in attitude toward *P* following his second communication if this behavior was greatly inconsistent with the hypothesis than if the act was in accord with *S*'s conception of *P*.

This prediction appears to differ from the results obtained in a number of studies in the area of "person perception." Typically, investigators have found a "primacy effect" in which the first-presented information concerning a person presumably gave rise to a general hypothesis about him, with later information then being assimilated to this hypothesis (Asch, 1952; Kelley, 1950; Luchins, 1957). For example, Kelley led one group of *S*s to believe that a speaker was a "warm" person, while other *S*s were induced to develop the hypothesis that he was "cold." Later, when the speaker delivered his talk, both groups of *S*s evaluated the speech in a manner consistent with their particular hypotheses about him. The present prediction, instead of assuming a similar assimilation tendency, if anything, argues for a "contrast (or recency) effect." It essentially maintains that *P*'s second action toward *S* is contrasted with the initial expectation, and as a consequence, influences *S*'s final attitude toward *P* to a greater extent than *P*'s first behavior.

The major difference between the present studies and the Asch-Kelley-Luchins investigations involves the discrepancy between *S*'s initial hypothesis and the information subsequently provided about *P*. It seems fair to assume that this later information in the Asch-Kelley-Luchins studies was not regarded as very different from *S*'s first concept of *P*, either because the information was somewhat ambiguous in nature or because it truly was not greatly inconsistent with the hypothesis. In summarizing Luchins' findings, Hovland (1958) stressed the fact that the *S*s did not consider the second set of information about the stimulus person to be incompatible with the first presented material, and he suggested that this is a limiting condition for the Luchins results. The suggestion is entirely in keeping with the present argument. There is reason to believe that if the *S*s had perceived the incompatibility between the two sets of material, a contrast would have occurred in which the later material would have been judged as being even more different from the first presented information than was actually the case (for example, as defined by independent judges [Berkowitz, 1960b]). The probable result of this contrast is a recency effect; the later information has a greater impact in affecting the final attitude toward *P* than does the first information. This essentially was the finding in the investigator's previous experiment (1960a), and presumably it arose from just this process: a contrast effect, in which *P*'s second act is evaluated as being extremely different from his first communication (and/or different from the initial hypothesis formulated concerning *P*), giving rise to the dominant influence of this later behavior upon *S*'s attitude toward *P*.

EXPERIMENT I

Method

Procedure. With some exceptions probably not relevant to our present purposes, the procedure of Experiment I was identical to that employed in the earlier investigation (Berkowitz, 1960a). Summarizing this procedure, two *S*s of the same sex were brought together ostensibly for a study of the formation of "first impressions." After receiving the instructions and a brief description of their task, they went into separate cubicles where they indicated their impressions of their partners on the first questionnaire (Q1). Three intercorrelated items in this questionnaire index *S*'s baseline level of friendliness toward *P*. However, the questionnaire purported to assess *S*'s judgments of *P*'s personality.

The *S*s then wrote two notes to each other at 5-min. intervals in which they described themselves. They were to use their partner's self-descriptions in evaluating him on the second questionnaire. *E* substituted previously constructed notes for the notes actually written in order to produce the initial hypothesis concerning *P*. The notes received by the *S*s in the Initial Hostility condition indicated that the partner had little regard for the University and most of its students, while *S*s in the No Initial Hostility condition were led to believe that the partner liked life at the University, and was fairly active in student affairs. We assume that the former *S*s expected *P* to be an unpleasant or hostile person as a result of these notes, while the latter *S*s expected him to be friendly. Following this, *S*s again indicated their impressions of their partners on a second form of the questionnaire (Q2). The same items used in Q1 were employed in testing the success of the initial hypothesis-creating manipulation.

E then collected Q2 and told the *S*s that they would now be given still more information about their partners: they would be told what their partners thought of them. *E* gave each *S* a previously completed Q2, saying that this was the partner's second rating of the *S*. Half of the *S*s in each of the above conditions (Unfavorable Evaluation) received an evaluation with uniformly unfavorable ratings, with the others (Favorable Evaluation) getting highly favorable ratings. After the *S*s had completed the third questionnaire (Q3) and an adjective check-list, the deceptions and experimental procedures were explained and the *S*s were dismissed with a pledge of secrecy from them.

Statistical analyses and dependent variables. The *S*s in each condition were differentiated according to whether their Q1–Q2 changes indicated either a decrease or an increase in unfriendliness toward *P* following his initial communications. *S*'s initial hypothesis concerning *P* may be inferred from these changes. A decrease in unfriendliness is taken as denoting an Expectation of Friendliness, while an increase in this unfavorable attitude defines an Expectation of Hostility. Within each of these groups, the *S*s were further subdivided in terms of whether their inferred hypotheses were strong or weak. They are said to have a Strong Expectation if their attitude changes after *P*'s initial communications coincided in direction with the changes displayed by the majority of *S*s (and therefore, presumably were in accord with the objective nature of the notes). The *S*s exhibiting Q1–Q2 attitude changes differing in direction from the majority of *S*s' responses to the particular set of notes are defined as having a Weak Expectation.[2]

The unfriendliness score involved in these attitude changes is based upon the responses to three intercorrelated items on the 13-item questionnaire forms. The items ask, "Would you admit him (her) into your circle of close friends?"; "Would you like this individual as a roommate?"; and "Would you admit him (her) into the social organization of greatest importance to yourself?" In each case the answers consist of a checkmark placed anywhere on a line ranging from Definitely yes to Definitely no, with the score being the distance away from the Friendly end of the scale in one-quarter inch units summed over the three items.

In addition to these questionnaire responses, the *S*s at the end of the experimental session were asked to describe *P* on an adjective check-list. Sixty-seven of the adjectives on this list had been shown in previous research to be particularly sensitive to changes of feeling on the evaluation dimension. Using the classifications of 63 undergraduate judges as a basis for dividing the terms into Good or Bad traits, *S*'s final level of unfriendliness toward *P* was assessed by adding the number of bad traits *P* was said to have to the number of good characteristics he was said not to possess. These scores were then "normalized" by means of a square-root transformation. For both the rating scale and check-list measures, our working assumption is that negative evaluations (or unfriendliness responses) are indicative of hostility.

Subjects. There are 160 *S*s (half of them women) in this first experiment, recruited from introductory psychology courses at the University of Wisconsin. These *S*s were distributed evenly, maintaining a one-to-one sex ratio, among the four experimentally-created conditions involving the combination of notes and evaluations from *P*. A preliminary analysis of variance of the data employing these classifications did not reveal any significant sex differences, and this variable is therefore ignored in the present report.

Results

Effectiveness of the initial hostility manipulation. There were no significant differences among the eight conditions in the mean level of unfriendliness toward *P* on the first questionnaire ($F = 1.38$ with 7 *df*). The frequencies of *S*s in each of the conditions, shown in Table 1, demonstrate that the initial notes generally were successful in establishing the intended attitude toward *P*; by far the greatest proportion of *S*s changed their attitudes from Q1 to Q2 in a direction consistent with these communications. (When the various evaluation conditions are grouped, since the *S*s had not received these evaluations at this stage, $\chi^2 = 21.04$; $p < .001$).

The same conclusion can be drawn upon inspection of the condition differences in mean changes at this time. If only the *S*s who had increased in unfriendliness toward *P* are considered, there was a significantly greater increase in the *S*s receiving the hostile notes than in the *S*s getting the friendlier communi-

---

[2] This reasoning assumes that all *S*s recognized, to some extent at least, the "objective" nature of *P*'s first communication. (Evidence consistent with this assumption is provided below.) Therefore, those whose feelings toward *P* changed in a direction inconsistent with this communication are regarded as being less certain of their initial hypothesis about *P*; they presumably had detected some evidence contradicting their initial belief.

TABLE 1

MEAN CHANGES IN QUESTIONNAIRE UNFRIENDLINESS TOWARD P

	Initial Hostility		No Initial Hostility	
	Unfavorable Evaluation (A)	Favorable Evaluation (B)	Unfavorable Evaluation (C)	Favorable Evaluation (D)
I Changes from Q1 to Q2				
*S*s who:				
(a) Increased in unfriendliness from Q1 to Q2	(32) 3.98^a	(24) 3.69	(14) 2.16	(13) 2.04
(b) Decreased in unfriendliness from Q1 to Q2	(8) −1.68	(16) −1.91	(26) −2.66	(27) −2.45
II Changes from Q2 to Q3				
Inferred initial hypothesis:	Strong-Hostility		Weak-Hostility	
(a)	(32) 0.92_c*	(24) -3.09_a	(14) 2.00_{cd}	(13) -0.84_b
Inferred initial hypothesis:	Weak-Friendly		Strong-Friendly	
(b)	(8) 2.95_{de}	(16) -1.24_b	(26) 3.26_e	0.90_c

[a] A positive score throughout signifies an increase in unfriendliness toward P.

* Cells containing the same subscript are not significantly different from each other (at the .05 level) by two-tailed t test.

cation (t_s for the Unfavorable Evaluation and Favorable Evaluation conditions are 4.14 and 3.34, respectively, both significant at $p < .01$).[3] The differences in attitude change were not as great in the case of the *S*s who decreased in hostility toward P from Q1 to Q2, but again the *S*s receiving the friendly notes tended to become somewhat friendlier (t_s = 1.75 and 1.23, respectively; the first difference is significant at the .10 level). These differences, it will be noted, are consistent with the assumption that attitude changes in keeping with the majority reaction to the messages denote a stronger expectation than changes deviating from the majority reaction. The degree of attitude change toward P is likely to be associated with the strength of the opinion S formed regarding P. Those who changed more may well have formed a more definite hypothesis.

Reactions to P's second acts. Generally, the Q2–Q3 attitude changes resulting from P's evaluations of S follow the same pattern as the results of the preceding study. The smallest shifts in hostility toward P, whether toward increased friendliness or unfriendliness, occurred in those conditions where the nature of the evaluations received from P was consistent with the inferred strong expectations developed from his earlier act. Thus, definitely expecting hostility from P, the *S*s in Condition A(a) showed a reliably *smaller* increase in unfriendliness after P's unfavorable evaluation of them than the *S*s in Condition C(b) who presumably strongly expected a friendlier act from him ($t = 4.59$; $p = .001$). They also exhibited a somewhat (but not significantly) smaller increase in unfriendliness toward P than the *S*s in Condition C(a) who had also received the unfavorable evaluation from P but who had only tentatively expected hostility from him ($t = 1.83$; $p = .10$). This change in Condition A(a), however, was significantly less than that for the *S*s in Condition A(b) who presumably had a weak hypothesis of friendliness from P ($t = 2.48$; $p = .02$).

Essentially the same pattern was displayed by the *S*s receiving the friendlier favorable evaluations from P. This action probably was consistent with the definite hypothesis held by the *S*s in Condition D(b), and they were much less likely to change toward friendliness following this favorable evaluation than the *S*s strongly expecting hostility (i.e., those in Condition B(a); $t = 4.18$; $p = .001$). The B(b) *S*s, only tentatively expecting friendli-

---

[3] The t tests (two tailed) conducted in Experiment I had their error terms based upon the "residual mean square" from a preliminary "among groups" analysis of variance. The tests of condition differences in Experiment II utilized the Duncan Multiple Range procedure (Duncan, 1955), which could be carried out because of the equal cell frequencies in this latter study.

TABLE 2

MEAN LEVEL OF UNFRIENDLINESS TOWARDS P ON THE ADJECTIVE CHECK LIST ADMINISTERED AT THE END OF EXPERIMENT I
(Square-Root Transformation)

	Initial Hostility		No Initial Hostility	
	Unfavorable Evaluation (A)	Favorable Evaluation (B)	Unfavorable Evaluation (C)	Favorable Evaluation (D)
*S*s who:				
(a) Increased in unfriendliness from Q1 to Q2	(32) 5.86	(24) 4.77	(14) 4.65	(13) 3.72
(b) Decreased in unfriendliness from Q1 to Q2	(8) 5.44	(16) 3.61	(26) 4.59	(27) 3.63

ness from P, apparently were more pleased by P's favorable evaluation of them than the former (D[b]) *S*s who supposedly had a strong hypothesis of friendliness from P ($t = 3.63$; $p = .01$). Similarly, *S*s in Condition D(a), tentatively expecting hostility from P, also became friendlier toward him upon receiving this pleasant act than did the D(b) *S*s ($t = 2.80$; $p = .01$).

To summarize, these results clearly are in accord with the predictions developed on the basis of the earlier investigation. *S*s receiving an act from P consistent with their presumed strongly held hypotheses about him showed relatively little change in attitude toward him. By and large, the greatest change in their feelings toward P was shown by *S*s who received acts that deviated from these definite hypotheses.

Cumulative effects of frustration-sequences. The data presented up to this point have referred to attitude changes. In actuality, of course, the *S*s' responses were in the form of absolute scores on the three-item rating scale. Inspection of the scores on the final questionnaire tends to confirm the *Frustration-Aggression* hypothesis regarding the cumulative effects of repeated frustrations. The *S*s in Condition A, who had received two unfriendly acts from P, had the highest level of hostility towards him by the end of the experiment, although these scores were not reliably greater than those in Condition C.

Similar results were found with the adjective check-list administered at the end of the experimental session, as can be seen in Table 2. The *S*s in Condition A(a), who had perceived two clear attacks upon them, described the attacker in a significantly more unfavorable manner than the *S*s perceiving only one definite attack. (Comparing the former condition with B(a), $t = 4.94$, $p < .01$, and with Conditions C(a) and C(b), it is 3.78, $p < .01$, and 4.38, $p < .01$, respectively.)

Nevertheless, it is clear from the previous analysis that this final level of hostility in Condition A(a) might not have increased much further with additional frustrations. S conceivably could learn to expect the frustrations as they are repeated so that their effect is lessened. However, this is not to suggest that a long series of frustrations would necessarily produce ever-diminishing reactions with the development of these expectations. The person may reach a "breaking point" after continually failing to reach his goals, for example, as McClelland suggests (1951), when he fears he cannot at all cope with the situation confronting him. His reaction then could well be a great surge of anger toward the frustrator.

EXPERIMENT II

The next experiment was designed to test the possibility that contrast effects in judgments of behavior are less likely to occur when S has a very general hypothesis about P. Conceivably, the difference between the foregoing results and the Asch-Kelley-Luchins findings might stem from differences in hypothesis-width. The *S*s in the latter investigations formed general impressions of the stimulus person, while *S*s' initial expectations concerning P in Experiment I might have been very restricted in scope, pertaining more directly to P's probable next act rather than to what he was like *in general*. If so, broadening this hypothesis to cover P as a person might result in an assimilation effect in which P's second act is interpreted in a manner consistent with the hypothesis.

Method

Procedure. Male and female college students who had volunteered to serve in an experiment on "personality factors in the ability to assess others' personalities" first participated in a group testing session in which they

TABLE 3

MEAN CHANGES IN LEVEL OF UNFRIENDLINESS TOWARDS *P* ON THE ADJECTIVE CHECK LISTS

Experiment II

	Hostile Expectation		Friendly Expectation	
	Hostile Note	Friendly Note	Hostile Note	Friendly Note
First Impression: Absolute score on CL-1	5.6	9.0	8.3	6.2
Change after expectation created	9.3[a]	8.3	−2.7	−2.7
Change after note	-1.3_b[*]	-5.4_a	4.3_c	0.44_{bc}

Note.—$N = 9$ in each condition.

[a] In all cases a positive score signifies an increase in unfriendliness toward *P*, while a negative score denotes a decrease in this attitude.

[*] Cells containing the same subscript are not significantly different from each other (at the .05 level) by Duncan Multiple Range Test.

responded to a personality inventory. Some time later two *S*s of the same sex, but who did not know each other, were scheduled for a given laboratory hour. When they arrived they were reminded of the ostensible purpose of the study and then ushered into separate rooms. There, each *S* indicated his first opinions of the other person (*P*) on a 33-item adjective check-list (CL-1), supposedly in order to demonstrate the accuracy of his first impressions, but actually to provide a baseline level of friendliness toward *P*.

*S*s were then given instructions designed to create initial expectations regarding *P*. After saying he would give them more information about their partner, *E* showed each *S* the partner's "personality profile" supposedly prepared by professional psychologists on the basis of the earlier test results. This "profile" consisted of graphic ratings on five bipolar traits. On four of these traits (for example, "introversion," "self-insight," etc.), the ratings indicated *P* was in the "normal" or "average" range. The fifth rating was varied with the experimental condition. Half of the *S*s were led to believe *P* was "extremely friendly" (Friendliness Expectation), while the others were shown that he was "very hostile" (Hostility Expectation). Once *E* was sure that the ratings were understood he gave *S* an alternate form of the adjective check-list (CL-2), saying he was to evaluate the partner "on the basis of all the information he now possessed." The scores obtained at this time were used to test the success of the manipulation of expectation.

In the last part of the experiment *S* and his partner were to write a note to each other in which they were to talk about themselves. The *S*s were told the note from *P* would provide them with further information about *P*. In actuality, however, they were given the previously prepared communications employed to manipulate attitudes in the preceding study. Half of the *S*s in each of the above conditions received the message, now combined on one card, used to increase unfriendliness (Hostility Note), and the remaining *S*s received the friendliness-increasing communication (Friendly Note).

After the notes were written, "exchanged" and read, each *S* rated his partner again on an adjective check-list (CL-3) that was composed of both alternate forms employed before. Four of the nine *S*s in each experimental condition were males.

Dependent variables. As in the first experiment, the unfriendliness score obtained from the check-list was computed by summing the number of unfavorable traits attributed to *P* with the number of good characteristics he was said not to possess. In testing the effect of the expectancy-creating statements, the CL-1 score for each *S* was subtracted from his CL-2 score. The final score, denoting the change in unfriendliness created by the communication from *P*, was obtained by considering only those adjectives in the final check-list that had been used in CL-2. The CL-2 score was subtracted from this score to yield the measure of the reaction to *P*'s behavior.

Results

The statistical findings obtained with the adjective check-list measures of unfriendliness are summarized in Table 3. Analysis of variance of the change scores following the expectation manipulation indicate that this treatment was highly successful. *S*s given the hostile expectation increased considerably in the unfavorableness of their descriptions of *P*, while those given the friendly expectation tended to develop a somewhat more favorable image of him. (This difference is significant at $p < .01$).

The reactions to the communication from *P*, shown in the last row of the table, generally were consistent with the results of the preceding study. The hostile note created the greatest increase in unfriendliness when it was sent to the *S*s having the friendly expectation, while there was the greatest decrease in unfriendliness following the friendly note when the message was received by *S*s presumably not expecting this type of behavior.

The implication of these findings is clear. It is unlikely that the previous results can be attributed to narrow hypotheses restricted to *P*'s probable next act. The contrast effects consistently obtained in these studies apparently arise because the second (or the to-be-evaluated) actions differ greatly from the actions expected by *S*.

DISCUSSION

The findings of these experiments can readily be understood as instances of judgmental phenomena. As is well known, when a person makes a judgment he essentially compares the stimulus with some standard, and a contrast effect occurs in this process to the extent that the stimulus differs from the standard (Berkowitz, 1960b). That is, *S* evaluates the difference as being greater than it really is (as defined, for example, by an observer not affected by the standard). In the present case, *S*'s expectation or hypothesis concerning *P* can serve as the standard against which *P*'s subsequent behavior is evaluated. When *P*'s action is altogether inconsistent with *S*'s hypothesis there apparently is an analogous contrast effect, and the action has a considerable impact upon *S*'s final attitude toward *P*. This judgmental analysis suggests, furthermore, that it may not be necessary to differentiate strong from weak hypotheses, a distinction, it will be recalled, introduced in Experiment I in connection with the correspondence between *S*'s and the majority reaction to *P*'s initial communication. For example, *S* was said to have only a tentative expectation concerning *P* if his attitude change toward *P* following this communication differed in direction from that exhibited by the majority of *S*s in his condition. Subsequent behavioral deviations from these presumably weak hypotheses did not have the same effect generally as deviations from supposedly stronger hypotheses. However, these perhaps unparsimonious interpretations concerning hypothesis strength can be avoided by making a somewhat different set of assumptions. Here we assume that the unfavorableness of *S*'s initial hypothesis about *P* can be coordinated directly to the changes in his questionnaire ratings of *P* from Q1 to Q2. Thus, the greater the change in an unfriendly direction, the more unfavorable is *S*'s hypothesis. If *P*'s second action be given an arbitrarily maximum score (say, 10) if it is aggressive in tone, and a score of 0 if it is friendly, the magnitude of the difference between this initial hypothesis and *P*'s subsequent behavior can then be computed. Extrapolating from judgmental research showing that the extent of the contrast effect is in direct ratio to the discrepancy between the stimulus and the standard, the greater the difference between the initial hypothesis and the later action, the greater should be *S*'s attitude change toward *P* following this action. Testing the prediction via the condition means shown in Table 1, the obtained correlation (tau) is +.71, which is significant at approximately the .03 level. The analogy therefore seems to hold. Unexpected behaviors may well be evaluated as being even more discrepant from the hypothesis than they really are, with the degree of the enhanced discrepancy a function of the degree of unexpectedness.

This reasoning has a direct bearing upon the hypotheses advanced by Dollard and his colleagues. Instead of supporting the Pastore-Bateson contention that it is necessary to distinguish among classes of frustrations because those that are unexpected are more likely than anticipated deprivations to produce aggressive reactions, the present findings can easily be incorporated into the Yale framework. According to Dollard, Doob, et al. (1939), the strength of the instigation to aggression is a positive function of the degree of frustration. We now can say that as a result of the contrast effect, unexpected frustrations are more likely to give rise to hostile responses because they probably are evaluated as being more severe than expected frustrations. In other words, there is a greater perceived interference with *S*'s goal attainment if the frustration is not expected than if it is anticipated because of the operation of judgmental processes upon the evaluation of the frustrating event.

The present analysis also questions the possibility raised by Miller and Campbell (1959) that there may be a "general primacy principle" which, at times, is masked by recency effects. The occurrence of primacy or recency may depend, in part at least, upon the perceived discrepancy between the first and later-presented information. The first material presumably gives rise to an hypothesis. In a sense, this expectation serves as the judgmental standard against which the later information is compared. If the later material appears consistent with the hypothesis (i.e., is close to the anchorage-expectation) it is incorporated into the hypothesis, and a primacy effect is said to take place; the first information has had greater weight than the later information in influencing the final evaluation of the stimulus person, situation, or issue. On the other hand, if the later material is seen as deviating considerably from the hypothesis (and assuming it is not rejected), there is a contrast effect exaggerating the difference between the two sets of information. In this case, the later material may have a greater impact upon *S*'s final attitude toward the stimulus object, not only because it is fresher in his memory, as Miller and Campbell point out, but because of the contrast-enhanced evaluation of the material.

SUMMARY

Two experiments were devised to test predictions developed on the basis of an earlier investigation (Berkowitz, 1960a). In this study employing pairs of *S*s it was found that the partner's act had its greatest effect on *S*'s final attitude toward him when it was inconsistent with the partner's earlier behavior. For example, if the second act was relatively aggressive in manner it provoked the greatest increase in unfriendliness toward the partner (*P*) when it had been preceded by friendlier behavior. The first experiment reported here, employing a procedure virtually identical to that used in the previous study, obtained essentially similar results.

The second experiment, yielding the same pattern of findings, showed that these results cannot be attributed to "narrow" hypotheses on the part of *S*s in the first study, restricted to *P*'s probable next act.

These contrast effects appear to differ from other findings in the area of "person perception". The discrepancies are interpreted in terms of a judgmental analysis. This judgmental analysis was also employed to make the present findings consistent with the frustration-aggression hypotheses advanced by Dollard, Doob, et al. (1939). The present paper suggests that unexpected frustrations produce a stronger aggressive reaction than anticipated frustrations because, as a result of the hypothesized contrast effect, the former probably are evaluated as being more severe.

REFERENCES

ASCH, S. *Social psychology*. New York: Prentice-Hall, 1952.

BATESON, G. The frustration-aggression hypothesis and culture. *Psychol. Rev.*, 1941, **48,** 350–355.

BERKOWITZ, L. Some factors affecting the reduction of overt hostility. *J. abnorm. soc. Psychol.*, 1960, **60,** 14–21. (a)

BERKOWITZ, L. The judgmental process in personality functioning. *Psychol. Rev.*, 1960, **67,** 130–142. (b)

DOLLARD, J., DOOB, L., MILLER, N., MOWRER, O., & SEARS, R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

DUNCAN, D. Multiple range and multiple F tests. *Biometrics*, 1955, **11,** 1–45.

HOVLAND, C. I. The role of primacy and recency in persuasive communication, In Eleanor E. Maccoby, T. Newcomb, & E. Hartley (Eds.) *Readings in social psychology*. New York: Holt, 1958.

KELLEY, H. H. The warm-cold variable in first impressions of persons. *J. Pers.*, 1950, **18,** 431–439.

LUCHINS, A. S. Primacy-recency in impression formation. In C. I. Hovland (Ed.) *The order of presentation in persuasion*. New Haven, Conn.: Yale Univer. Press, 1957. Pp. 33–75.

MASLOW, A. H. Deprivation, threat, and frustration. *Psychol. Rev.*, 1941, **48,** 364–366.

MCCLELLAND, D. C., & APICELLA, F. A functional classification of verbal reactions to experimentally induced failure. *J. abnorm. soc. Psychol.*, 1945, **40,** 376–390.

MCCLELLAND, D. C. *Personality*. New York: Dryden, 1951.

MILLER, N. E. The frustration-aggression hypothesis. *Psychol. Rev.*, 1941, **48,** 337–342.

MILLER, N., & CAMPBELL, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *J. abnorm. soc. Psychol.*, 1959, **59,** 1–9.

OTIS, N., & MCCANDLESS, B. Responses to repeated frustrations of young children differentiated according to need area. *J. abnorm. soc. Psychol.*, 1955, **50,** 349–353.

PASTORE, N. The role of arbitrariness in the frustration-aggression hypothesis. *J. abnorm. soc. Psychol.*, 1952, **47,** 728–731.

(Early publication received November 5, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

CRITIQUE AND NOTES

FRUSTRATION, AGGRESSION, AND MURDER[1]

STUART PALMER
University of New Hampshire

IN THE study reported here, 51 male murderers and their 51 nearest-age brothers were investigated with respect to frustration in infancy, childhood, and adolescence as an influencing variable (cf. Dollard, Doob, Miller, Mowrer, Sears, Ford, Hovland, & Sollenberger, 1939). Frustration was defined as lack of satisfaction of an innate or learned need. A murderer was defined as a person who has been legally convicted of first or of second degree murder or of first degree (non-negligent) manslaughter. The central hypothesis was: There is a significant positive relationship between the amount of frustration experienced by persons in infancy, childhood, and adolescence, on the one hand, and whether or not they later commit murder, on the other hand. Aggression release during the preadult years was considered a major control variable.

METHOD

The records of all males serving prison sentences for murder or first degree manslaughter in Massachusetts, New Hampshire, Maine, and Rhode Island during the fall of 1957 were investigated. In order to be used as cases in this study, it was necessary that, as well as having brothers, the murderers have mothers who were available for interviewing. Of 254 imprisoned men, 76 met the following Criteria: 1. under 51 years of age; 2. court records indicated that the subject was believed to have physically committed the murder; 3. at least one brother who had attained the age the murderer was when he committed murder; 4. mother living; 5. mother able to answer questions in English; and 6. mother residing within 150 miles of Durham, New Hampshire. Twenty-five of the 76 mothers were not available for interviewing, leaving 51 *S*s in the murderer group. Where a murderer had more than one brother who fitted Criterion 3 above, the brother nearest the murderer in age was chosen, to yield a control group of the same size. None of these control brothers was known to have committed murder. Use of the brothers as control *S*s made it possible to hold relatively constant such factors as age, sex, social class of parents, religion, and the like.

Some data, such as intelligence quotients, were obtained from the prison records of the murderers. Most were obtained through interviewing the mothers with the use of a questionnaire composed of 143 items, a majority of which were close-ended. Each mother completed the questionnaire twice, once for the murderer son, once for the nonmurderer. The order of questioning about sons was reversed with alternate mothers.

Indices of Physical Frustration, Psychological Frustration, General Frustration, Socially Acceptable Aggression Release, Socially Unacceptable Aggression Release, and General Aggression Release were developed. In the Index of Physical Frustration, the *S* was allotted one point for the following:

1. A difficult birth
2. Serious effects of forceps at birth
3. Each serious operation
4. Each serious illness
5. Each serious accident
6. Each serious beating by someone other than a parent or step-parent
7. One or more serious beatings by the mother; and
8. One or more serious beatings by the father or step-father; with respect to this index as well as to those below, words such as difficult and serious were operationally defined

In the Index of Psychological Frustration, one point was allotted the *S* for the following:

1. Each serious, physical defect that could give rise to social disapproval
2. Very rigid time schedule adhered to by the mothers in caring for the child during his first two years
3. Bottle-feeding during some portion of the first two months of life
4. Fixed schedule followed by mother in feeding the child during the first two months
5. Final weaning started before age 12 months
6. Toilet-training begun before age 12 months
7. Sexual self-play stopped by mother before age 12 months
8. Mother responded to child very often by crying during first five years of his life
9. Mother responded to child very often by isolating herself during first five years of his life
10. Each trauma due to natural events or other individuals, excluding beatings
11. Great difficulty learning to read or never learned

---

[1] Supported in part by four grants from the Central University Research Fund of the University of New Hampshire.

12. Persistent bed wetting
13. Persistent stuttering; and
14. Spent a majority of time playing alone at age five

An *S*'s score on the Index of General Frustration was found by adding together his scores on the Index of Physical Frustration and the Index of Psychological Frustration.

In the Index of Socially Acceptable Aggression Release, one point was allotted the *S* for the following:

1. Verbal loudness at age five classified as loud or average
2. Anger short of temper tantrums at age five once a week or more
3. Reasonable verbal release at age five which occurred half or more of the times when angry
4. Physical fighting at age five once a week or more
5. "Talking back" to the mother at age five occasionally or more
6. Dirtying clothes at age five, half or more of the times when "dressed up"
7. Participation in athletics at age 17 occasionally or more
8. Hunting at age 17 occasionally or more; and
9. Mildly antagonizing others by verbal means at age 17 occasionally or more

In the Index of Socially Unacceptable Aggression Release, one point was allotted the *S* for the following:

1. Lying to the mother at age five occasionally or more
2. Stealing at age five occasionally or more
3. Swearing at age five occasionally or more
4. Temper tantrums once or more during childhood
5. Deliberately and seriously hurting animals once or more during childhood or adolescence
6. Fist-fighting once or more at age 17
7. Violent aggression other than fist-fighting when intoxicated, before age 20

Whether a behavioral form indicated socially acceptable or unacceptable aggression release was determined arbitrarily by the investigator. An *S*'s score on the Index of General Aggression Release was found by summing his scores on the Index of Socially Acceptable Aggression Release and the Index of Socially Unacceptable Aggression Release.

RESULTS

As Table 1 shows, significantly more murderers than control brothers scored high on the Indices of Physical, Psychological, and General Frustration. Further, when General Frustration scores within pairs of murderers and brothers were compared, it was found that in 42 pairs (82%) the murderer had

TABLE 1

SCORES OF MURDERERS AND CONTROL BROTHERS ON INDICES OF FRUSTRATION AND AGGRESSION

Score	Physical Frustration (No. of)		Psychological Frustration (No. of)		General Frustration (No. of)		Acceptable Aggression (No. of)		Unacceptable Aggression (No. of)		General Aggression (No. of)	
	M	B	M	B	M	B	M	B	M	B	M	B
0	4	13	2	4	0	2	5	2	22	35	5	1
1	5	13	2	11	1	4	10	5	12	12	6	3
2	8	12	6	13	3	10	9	12	6	3	7	15
3	5	7	10	11	2	7	16	11	8	1	7	7
4	3	5	7	4	4	5	7	4	2		9	5
5	6	0	5	5	7	8	3	4	1		7	4
6	3	1	6	2	2	6	1	8			2	10
7	7		6	1	2	3		5			4	2
8	6		2		5	3					2	3
9	3		1		5	2					2	0
10	0		4		1	1						1
11	1				1							
12–21					18							
$\bar{X}$	4.5	1.7	4.7	2.5	9.2	4.2	2.4	3.5	1.2	0.4	3.7	4.0
Chi-square	33.10		22.10		30.36		13.13		12.30		13.76	
P	$<.01$		$<.01$		$<.01$		$<.05$		$<.01$		$<.05$	
Scores combined for chi-square	5–11		0–1; 7–10		0–2; 3–4; 5–6; 7–8; 9–10; 11–21		0–1; 5–7		3–5		0–1; 7–10	

Note.—M designates murders; B designates control brothers. In each column, $N = 51$.

a higher score than his brother. In five pairs the scores were equal, and in four, the brother's score was higher.

The data in Table 1 also show that a significantly greater number of control brothers than of murderers had high scores on the Index of Acceptable Aggression Release. Conversely, a significantly greater number of murderers than of brothers had high scores on the Index of Unacceptable Aggression Release. On the Index of General Aggression Release, scores were approximately the same for the murderers as for the brothers (in spite of the fact that the Chi square value was significant at the .05 level).

Although detailed data are not included here, the relationship between general frustration and murder was investigated while holding general aggression release approximately constant. This was done by dichotomizing the total group of 102 subjects into two equal sub-groups; one composed of 51 murderers and brothers who used general aggression releases extensively (score of 4 or more on the Index of General Aggression Release), and the other composed of 51 murderers and brothers who used general aggression releases infrequently (score of three or less). Frustration was positively associated with murder whether aggression releases were used extensively or infrequently. However, the association was statistically significant only when aggression releases were employed extensively.

DISCUSSION

Clearly, there are three especially major drawbacks in this study. First, the mothers may not have reported the facts correctly. Second, the group of murderers is probably not representative of any larger murderer population. Third, the indices of frustration and aggression release are crude; they do not take into account the fact that one incident of frustration or aggression release may not equal the weight of another such incident. However, as an exploratory study where more satisfactory data are difficult to obtain, it indicates the possibility of a functional relationship between preadult frustration and later presence or absence of murder.

REFERENCE

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., SEARS, R. R., FORD, C. S., HOVLAND, C. I., & SOLLENBERGER, R. T., *Frustration and aggression*. New Haven: Yale Univer., 1939.

(Received June 11, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SOME PHYSIOLOGICAL CORRELATES OF RORSCHACH BODY IMAGE PRODUCTIONS[1]

A. D. DAVIS[2]
University of Texas

RECENT investigations have given rise to the contention that some aspect of personality organization reflected in Rorschach percepts is involved in the selection of psychosomatic symptomatology (Cleveland & Fisher, 1954; Fisher & Cleveland, 1955, 1956, 1957). This personality variable is concerned with the individual's perception of his body exterior as having (or lacking) protective characteristics. The relative strength of this "body-image perception," according to Fisher and Cleveland (1957), can be quantified as a "barrier score" by an appropriate scoring procedure. This barrier score has been related to the projective performance of clinical groups manifesting specific psychosomatic symptom complexes, and it has been argued that these bodily perceptions predispose normal individuals to symptoms at either an "interior" or "exterior" bodily site through heightened physiological response tendencies (Fisher & Cleveland, 1955, 1957).

This formulation suggests that the greater number of Rorschach responses which emphasize "barrier" (impermeable, enclosing, or otherwise protective) surface qualities, the greater is the physiological reactivity toward the periphery of the body (skin and musculature). Conversely, the relative absence of these productions is related to increased physiological reactivity at an interior body site. The present study was formulated as a means of investigating this hypothesized relationship in normal subjects.

---

[1] This article is based on a dissertation submitted to the faculty of the department of psychology of the University of Texas in partial fulfillment of the requirements for the PhD degree. The author is indebted to Philip Worchel, Seymour Fisher, and Sidney E. Cleveland for their support, aid, and supervision in this study.

[2] Now at the Department of Psychology, Western State Hospital, Ft. Steilacoom, Washington.

METHOD

Subjects

Group Rorschachs were administered to three sophomore psychology classes totaling 280 students at the University of Houston. Of these *S*s, only the 146 males were utilized because of the nature of the physiological recording. From *S*s having more than the median number of barrier responses, 25 were randomly selected and designated as the "high barrier" group (Group I); of *S*s with less than the median number of barrier responses, 25 were randomly selected and designated as the "low barrier" group (Group II).

Apparatus and Physiological Measures

Physiological indices were obtained by the following methods and apparatus:

1. *Electromyogram* (EMG). Recordings of muscle action potential were obtained from two EMG electrodes fixed by a head band placed two inches apart over the frontalis muscles of the left eye. The muscle potentials were fed into an electronic current integrator ter the design of Jacobson (1951) and this integral recorded on a standard Grass III-C, 7 channel, ink-writing oscillograph.

2. *Skin Resistance* (GSR). A measure of skin resistance was obtained from two 2-inch sponge electrodes placed on the dorsal and palmar surface of the left hand. The basic circuit of the GSR unit utilized for recording is reported by Davis and Porter (1931).

3. *Blood Pressure.* Systolic (SBP) and diastolic (DBP) blood pressures were obtained by the cuff and auscultatory method from the right arm. Mean blood pressure (MBP) was computed as the average of the systolic and diastolic pressures.

4. *Ballistocardiograph Variables* (BCG). Four ballistocardiograph indices were utilized after the fashion of Funkenstein, King, & Drolette (1954): (*a*) An absolute count of pulse rate (P). (*b*) The mean height of the BCG-IJ wave was determined over the arbitrary time span of the IJ waves utilized to define stroke volume (IJ). The IJ wave-form characteristic in the BCG record is the most conspicuous characteristic of the record and it is generally agreed to be the headward ballistic reaction of the body to ventricular systole (Dock, Mendelbaum, & Mendelbaum, 1953). This variable has been empirically related to ventricular stroke volume by Starr, Schnabel, Mayock, & Austin, 1952). (*c*) The average height of the IJ unit was multiplied by the pulse rate over unit time (IJ × P). This measure was empirically related to cardiac output by Nickerson, Warren, & Brannon, 1947). As outlined by Funkenstein (1954), these variables were utilized without the empirically derived constants of Starr and Nickerson (1952) as purely relative physiological indices with no interpretation of absolute physiological meaning or value intended. (*d*) The mean blood pressure divided by the average height of the IJ waves multiplied by pulse defined peripheral resistance (MBP/IJ × P). This measure has similarly been empirically related to the total peripheral resistance of the cardiovascular system.

The instrument used to obtain these records was a critically damped accelerometer consisting of a two-gram weight pendulum which was suspended from a magnetic phonograph cartridge mounted in a block of wood 18 × 4 × 2 inches. The pendulum weight was immersed in a glycerin bath which prevented reverberation of the pendulum, thus giving clean records of the body ballistics. The accelerometer was designed to give identical wave form characteristics to that of the Virgil Hill Electromagnetic BCG (Dock, Mendelbaum, & Mendelbaum, 1953). The apparatus was placed across the legs at the ankles and recorded on the Grass oscillograph.

From these measures, two groups of three variables each were assigned to exterior and interior classifications. The rationale for this division was dependent upon the bodily site at which the greatest variation of the measures occurred. The exterior variables and the physiological process to which they are related were as follows: (*a*) muscle potential (EMG), (*b*) skin resistance (GSR), and (*c*) peripheral resistance (MBP/IJ×P).

The interior measures included (*a*) pulse rate (P), (*b*) stroke volume (IJ), and (*c*) cardiac output (IJ×P). Systolic and diastolic blood pressures (SBP) and (DBP) were obtained in order to compute the mean blood pressure (MBP).

According to the formulation, Group I (high barrier) individuals should show greater responsivity in the exterior group of variables, and Group II should reflect relatively greater physiological response on the measures designated as interior.

Procedure

Each *S* was instructed to lie down on a one-half inch horsehair pad on the floor. A sandbag was placed beneath his heels to lock his knees for the ballistocardiograph recording, and the remainder of the equipment was affixed while the investigator casually engaged him in conversation about neutral topics. S was then instructed to relax and rest while the investigator obtained "recordings of his heart." An initial rest period of 20 min. was given in an attempt to establish conditions which were as basal as possible for resting-level measures. During the last 10 min. of the rest period, GSR readings were recorded, and the blood pressure was sampled three times. At the end of the rest period, a 2-min. sample recording of EMG and BCG was obtained and utilized as rest level measures. At this time, another sample of GSR and blood pressure was taken and averaged with the initial measures as an estimate of resting level.

Six arithmetic problems from the Wechsler Intelligence scale were read from cards. S was given only 30 sec. for each solution. The first two answers were accepted as correct and the last four as incorrect, regardless of their nature. The problems were read readily, and the investigator's standard comments after each "failure" were designed to be increasingly more depreciating. The GSR reading was noted after presentation of each problem, and the maximum fall in resistance during the problem period was considered as the maximum stress response on that variable. After presentation of the last problem, blood pressure was immediately taken, and a sample recording of BCG and EMG was obtained. These measures were utilized as the poststress (or response to stress) measures.

RESULTS

The scores of the combined groups on each variable were linearly transformed into distributions having a mean of 50 and a standard deviation of 10, making it possible to compare the individual's performance on any variable with his performance on any other relative to the sample mean of the variable. As a means of evaluating the differences in the groups' resting levels, t and F tests of significance were performed between groups on each of the variables in the pre-stress condition. The significant data are presented in Table 1. Marked differences occurred between the two groups on their resting measures. Group I was significantly higher on one exterior measure (EMG) at the .02 level with a one-tailed test. The Group II resting response level, as predicted, tended to be higher than that of Group I on the interior measure of P. However, the outcomes for IJ and IJ × P were reversed with Group I having the higher mean. Interestingly, the observed blood pressure measure, SBP and MBP, approached significance in favor of Group II and appeared to have greater variance than any other Group II performance. Similarly, Group I variance on the blood pressure measures was among the smallest of that group's performances. Inasmuch as the standard deviation for the distribution of the groups combined is equal to 10, discrepancies from this value are some relative indication of the lability of the physiological variable for the group.

TABLE 1

TRANSFORMED PRESTRESS MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF DIFFERENCES

Physiological Index	Group I		Group II		Mean Diff.	t	P (1-tailed)	F
	Mean	*SD*	Mean	*SD*				
Exterior								
EMG	53.06	9.71	46.94	9.50	6.12	2.26	.02	—
GSR	51.42	7.57	48.37	9.64	—	—	—	—
MBP/IJ × P	49.22	11.10	51.00	9.10	—	—	—	—
Interior								
P	47.97	10.78	52.03	8.90	4.06	1.45	.08	—
IJ	52.56	11.06	47.18	9.15	5.38	1.87	.045	—
IJ × P	51.11	11.73	48.41	7.18	—	—	—	2.66
Blood Pressure								
SBP	47.95	8.51	52.45	10.37	4.50	1.68	.09	—
DBP	48.92	8.85	51.08	11.07	—	—	—	—
MBP	47.50	8.46	52.49	10.93	4.99	1.81	.05	—

To make comparisons between the groups on their relative physiological reactivity to stress, pre- and postdifference scores were corrected for differences in initial resting level for a regression analysis. The procedure was recommended by Lacey (1956), who pointed out that the high correlation between resting and stress levels predisposes percentage or absolute-change data toward reflecting initial differences. The result of this treatment yielded a distribution of "Autonomic Lability Scores" free of dependence on base level and expressed in units which make it possible to compare relative responsivity on a number of different physiological variables.

Table 2 summarizes the data for the Autonomic Lability Scores. It is apparent that Group I responded to the stress situation in terms of the exterior variables with greater EMG, GSR, and MBP/IJ×P values than those of Group II. Although the GSR difference was not significant, the difference was in the predicted direction. With a one-tailed test, EMG and MBP/IJ×P may be considered as significant at the .01 and .06 levels

TABLE 2

AUTONOMIC LABILITY SCORE MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF DIFFERENCES BETWEEN GROUPS

Physiological Index	Group I		Group II		Mean Diff.	*t*	*P* (1-tailed)	*F*
	Mean	*SD*	Mean	*SD*				
Exterior								
EMG	53.39	9.34	46.81	9.60	6.58	2.46	.01	—
GSR	51.15	7.76	49.49	5.74	1.66	.86	.20	—
MBP/IJ × P	51.66	8.58	47.49	9.65	4.17	1.62	.06	—
Interior								
P	47.37	8.60	52.62	9.72	5.25	2.03	.02	—
IJ	47.87	7.56	53.08	11.05	5.21	2.17	.03[a]	2.13
IJ × P	47.81	6.74	52.25	8.30	4.44	2.07	.02	—
Blood Pressure								
SBP	45.89	9.20	53.73	9.26	7.64	3.00	.005	—
DBP	48.66	10.84	51.27	8.54	2.61	.94	.17	—
MBP	46.30	10.15	53.98	7.72	7.68	3.01	.005	—

[a] Two-tailed test due to significant *F* ratio.

TABLE 3

RANK ORDER OF RESPONSIVITY AND COEFFICIENTS OF THE TWO GROUPS ON EACH OF THE EXTERIOR AND INTERIOR VARIABLE IN PRE- AND POSTSTRESS AND AUTONOMIC LABILITY MEASURES

Physiological Index	Prestress Ranks		Poststress Ranks		Prepost Autonomic Lability Scores Ranks	
	Group I	Group II	Group I	Group II	Group I	Group II
Exterior						
EMG	1	6	1	6	1	6
GSR	3	4	2	5	3	4
MBP/IJ × P	5	2	4	4	2	5
Interior						
P	6	1	6	1	6	2
IJ	2	5	3	3	4	1
IJ × P	4	3	5	2	5	3
rho:*	−1.00		−.86		−.83	
Standard error	—		.111		.133	

* rho of .829 needed at .05 level of significance; .943 needed at .01 level.

respectively. Conversely, Group II responded in terms of the interior variables with P (t of 2.03), IJ (t of 2.17), and IJ×P (2.07) all significant beyond the .02 level. A one-tailed test was not reported on the IJ variable because of the significant difference in variances. The blood pressure measures, SBP and MBP, as in the resting level, were greater (.01) in Group II.

With the raw data transformed in such a way as to permit the relative order of responsivity across variables to be determined, the mean rank of each group on each variable was found and rank order (rho) correlations computed. These ranks and their rho coefficients for the prestress, poststress, and pre- and poststress regression analyses (Autonomic Lability Scores) are presented in Table 3. As the table suggests, there was a marked contrast in the form of a high inverse relationship. There was a perfect negative correlation between the order of Group I and Group II in the resting level; for the poststress and difference analyses, the rhos of −.86 and −.83 respectively were both significant beyond the .05 level.

DISCUSSION

The results indicate that the differences at resting level between the groups are not completely in accord with the hypotheses. For Group I (high barrier), however, the EMG measure is greater than that of the Group II (low barrier) measures. Similarly, for Group II, P is greater than in Group I with no evidence of differences between the groups on GSR, MBP/IJ×P, or IJ×P. It will be noticed that IJ×P (cardiac output) enters into both of the nonsignificant cardiac variables. It appears from inspection of the underlying variables (P and IJ) that the cardiac output, although not different, may be achieved in different ways during the resting measures. In Group I, it seems to be primarily a function of significantly higher stroke volume (IJ) and slow pulse; whereas in Group II, the cardiac output appears to result from low stroke volume and a significantly more rapid pulse. Additionally, the pre- to poststress means suggest that all cardiac activity in Group I decreases from rest to stress. This is undoubtedly related to the increase in this group's peripheral resistance, which would tend to impede any cardiac output. Nevertheless, cardiac activity in Group II seems excessive under stress in comparison to the increase shown by Group I *S*s. On the other hand, the cardiac organization of Group I appears to becomes more efficient under stress, possibly through a slowing of the pulse, allowing greater ventricular filling before systole. These observations reflect highly significant differences in the organization of the two groups' cardiovascular functioning. Such characteristic differences in cardiovascular organization across individuals were similarly reported by Lacey (1950, 1953) with *S*s under stress.

To summarize the resulting physiological hierarchies of response of the two groups under stress: Group I (high barrier) reacts in terms of increased muscle potential and peripheral vasoconstriction, a decreased skin resistance, increased stroke volume of the heart, increased cardiac output, and increased pulse rate in that order. The Group II (low barrier) hierarchy of response under stress consists of an increase of stroke-volume, pulse rate, and cardiac output, and a decrease in GSR, increased peripheral resistance, and muscle potential in that order. These hierarchies clearly demonstrate a primacy of response in the exterior indicies for the high barrier group and for the interior indices in the low barrier group when both groups are under stress. These data tend to lend support for the hypotheses of this study if the hypotheses are restricted to considerations of the stress data. The resting level data do not support the contention that these physiological predispositions are characteristic of the two groups under basal conditions.

SUMMARY

The relationship between surface characteristics of Rorschach responses and physiological measures was investigated in order to test the hypothesis that the projected perception of one's body is related to the patterning of physiological responses. Nine physiological measures were obtained under rest and experimentally induced stress from 50 *S*s selected at random from a larger sample of 146 normal male subjects. Physiological indices related to bodily exterior (muscle potential, skin resistance, cardiovascular peripheral resistance) and bodily interior (pulse rate, stroke volume, cardiac output) physiological responsivity were abstracted from the nine measures. The results indicate that the tendency to give impermeable or protective characteristics to Rorschach percepts is related to heightened physiological reactivity at an exterior body site and that the absence of these characteristics in Rorschach productions is related to heightened physiological response on interior indices of physiological response under stress.

REFERENCES

CLEVELAND, S. E., & FISHER, S. Behavior and unconscious fantasies of patients with rheumatoid arthritis. *Psychosom. Med.*, 1954, **15**, 327–334.

DAVIS, R. C., & PORTER, J. M., JR. A measuring device for the galvanic skin reflex. *J. gen. Psychol.*, 1931, **5**, 115–120.

DOCK, W., MENDELBAUM, H., & MENDELBAUM, R. A. *Ballistocardiography.* St. Louis: C. V. Mosby, 1953.

FISHER, S., & CLEVELAND, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.*, 1955, **69**, (17, Whole No. 402).

FISHER, S., & CLEVELAND, S. E. Relationship of body image to site of cancer. *Psychosom. Med.*, 1956, **18**, 304–310.

FISHER, S., & CLEVELAND, S. E. An approach to physiological reactivity in terms of a body-image schema. *Psychol. Rev.*, 1957, **64**, 26–38.

FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. The direction of anger during a laboratory stress-inducing situation. *Psychosom. Med.*, 1954, **15**, 404–413.

JACOBSON, E. Muscular tension and the estimation of effort. *Amer. J. Psychol.*, 1951, **64**, 112–117.

LACEY, J. L. Individual differences in somatic response patterns. *J. comp. physiol. Psychol.*, 1950, **43**, 338–350.

LACEY, J. The evaluation of autonomic responses. Toward a general solution. *Ann. NY Acad. Sci.*, 1956, **67**, 123–164.

LACEY, J. I., BATEMAN, DOROTHY E., & VAN LEHN, RUTH. Autonomic response specificity. *Psychosom. Med.*, 1953, **15**, 8–21.

NICKERSON, J. L., WARREN, J. V., & BRANNON, E. S. Cardiac output in man: Studies with the low frequency, critically-damped ballistocardiograph and method of right atrial catheterization. *J. clin. Invest.*, 1947, **26**, 1–10.

STARR, I., SCHNABEL, T. G., JR., MAYOCK, R. L., & AUSTIN, J. H. Relation of aortic pulse pressure curves both to left ventricular stroke volume and to output at every instant of systole. *Fed. Proc.*, 1952, **11**, 152.

(Received October 7, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

MANIFEST ANXIETY AND PRESTIMULUS CONDUCTANCE LEVELS

GERARD J. McDONNELL AND JOHN A. CARPENTER
Yale University

NUMEROUS studies reporting correlations between conductance or palmar sweating changes and changes in anxiety stimuli or self-report of tension have encouraged the view that conductance level, from which the changes have been noted, represents the anxiety or tension[1] level of an individual. But studies attempting to determine the function relating conductance levels and manifest anxiety are rare. Hence, it is a dubious procedure to infer that one individual is more anxious than another on the evidence that his conductance levels are higher. Moreover, lacking knowledge of the function between anxiety and conductance level, it is, strictly speaking, improper to interpret changes in level as unequivocal evidence of changes in anxiety.

The following study was designed to clarify this problem. Specifically it was planned to obtain data on two questions:

1. If Individual A has the same prestimulus level as B, do the two have the same manifest anxiety level?
2. If A has a different level than B, does the one with the higher conductance also have the higher anxiety level?

METHOD

Thirty-two men, all graduate students, and eight women, students or students' wives, aged 21 to 28, served as paid subjects (*S*s).

While *S*s played cards, electrodes were fastened to the plantar surfaces of both feet. Essentially simultaneous recordings were taken from groups of eight *S*s by a method described elsewhere (McDonnell & Carpenter, 1959). Recordings lasted for 15 min. for each of the two sessions (spaced one week apart). The individual's mean conductance level for both sessions was taken as his prestimulus conductance level. Manifest anxiety level was obtained by administration of the Mandler-Sarason General Anxiety scale at the second session.

RESULTS AND DISCUSSION

The questions prompting this inquiry were (*a*) if A and B have the same conductance level, do they report the same anxiety, and (*b*) if B has a higher level than A, does he also report greater anxiety?

Answers are obtained by referring to the relation between anxiety and conductance shown in Fig. 1. Each plotted point represents an average of the conductance level of those *S*s whose anxiety scores were found in a given anxiety interval. The correlation ratio was .722 ($P < .01$ with 39 *df*).

Thus, it appears that if A and B both have high conductance levels, they are apt to report moderate anxiety; but if they have low conductance levels, *S* could report either high or low anxiety. A similar difficulty obtains in predicting differences in anxiety from the fact that B has a higher conductance than A. The higher conductance signifies higher anxiety only when both *S*s report relatively low anxiety. If both report high anxiety, the higher conductance would be expected in the *S* with lower anxiety.

The evidence that low conductance may be associated with high anxiety corresponds to two previous observations. Van der Valk and Groen (1950) found that *S*s with diseases supposedly symptomatic of high tension (ulcerative colitis, peptic ulcer, diabetes, etc.) showed lower conductance levels than a group of nondiseased normals. Similarly, Mowrer, Light, Luria, & Seleny (1953) reported an overtly high anxious patient who had "phenomenally" low levels of palmar sweating. The idea that similar conductance levels can characterize *S*s with different anxiety levels is supported by the same author, who noted that individuals differing widely in tension level showed nearly identical sweating levels.

Greater awareness of anxiety may be accompanied at first by greater autonomic activity. Thereafter, some process, possibly an extension of homeostatic restraint as conceived by Lacey (1956), intervenes to modulate sympathetic effects. The result is a lowering of conductance levels.

Although this study only furnishes evidence on the meaning of conductance *levels*, the finding of a curvilinear function suggests caution in interpreting conductance *changes*. Depending on *S*'s level of anxiety, a decrease in conductance could signify an increase in anxiety for one *S* and just the reverse for another. Illustratively, Mowrer found with some patients that a decline in sweating was accompanied by an increase in reported tension. Similarly, as other investigators (Van der Valk & Groen, 1950) have interpreted their results, an increase in conductance could, in some *S*s, mark a decrease in tension.

---

[1] The study assumes, without proof, an equivalence between "anxiety" and "tension." "Anxiety" will be used hereafter for both.

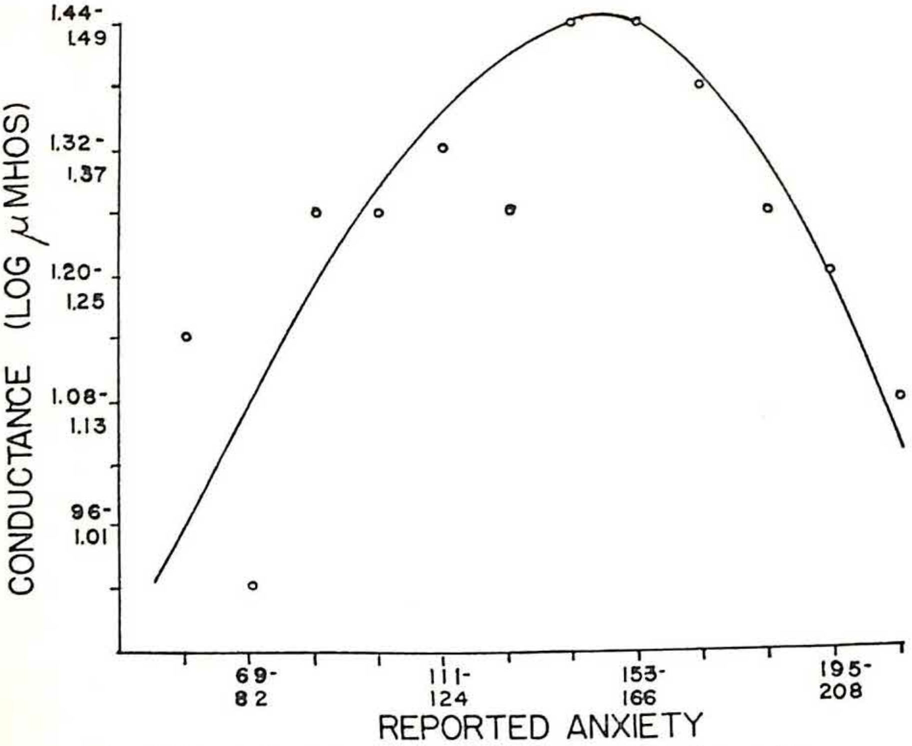


FIG. 1. The relation between skin conductance and manifest anxiety.

The results of this study require two qualifications: size and nature of sample, and the validity of the anxiety scale.

Sample size is important where correlation ratios are concerned, because small frequencies in the grouping intervals exaggerate the coefficient. Although care was taken to use gross categories, the number of *S*s at the extremes of the distribution were nevertheless small. Again, the twin circumstances of being voluntary and paid might have selectively influenced both conductance and anxiety score. However, the absence of satisfactory norms for conductance or anxiety levels makes identification of such influences impossible. Including results of both sexes seemed warranted for two reasons: females resembled the males in range of both anxiety scores and conductance levels, and they exhibited the same curvilinear relation between the two variables.

While reliability of the Anxiety scale is satisfactory, validity is not so well established (Mandler & Sarason, 1952). The lower anxious group was found here to be more heterogeneous in conductance level than the higher anxious group, possibly indicating that such low scores include both genuinely low anxious *S*s and *S*s who are denying or suppressing anxiety. It is equally possible that the greater variability of this group results from (*a*) weakness of discrimination in the Anxiety scale or (*b*) lack of correspondence between anxiety and conductance at the low end of the anxiety continuum. The data do not permit a decision between these alternatives.

SUMMARY

This study investigated the functional relationship between conductance levels and an index of manifest anxiety. A curvilinear relationship was found and discussed in relation to the use of conductance level as an index of emotionality.

REFERENCES

LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. NY Acad. Sci.* 1956, **67,** 123–164.

McDONNELL, G. J., & CARPENTER, J. A. Anxiety, skin conductance and alcohol. *Quart. J. Stud. Alcohol,* 1959, **20,** 38–52.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47,** 166–173.

MOWRER, O. H., LIGHT, B. H., LURIA, Z., & SELENY, M. P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy*. New York: Ronald, 1953.

VAN DER VALK, J. M., & GROEN, J. Electrical resistance of the skin during induced emotional stress. *Psychosom. Med.*, 1950, **12,** 303–314.

(Received July 23, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

A PSYCHOLOGICAL FACTOR APPARENTLY ASSOCIATED WITH NEOPLASTIC DISEASE[1]

LAWRENCE LeSHAN
The Institute of Applied Biology, New York

AND MARVIN REZNIKOFF
The Institute of Living, Hartford, Connecticut

An investigation of 250 patients with malignant disease (LeShan, 1956) indicated a pattern in 62% of this group and in 10% of the controls which included a psychologically traumatic event early in life and a marked amount of disturbance in the individual's ability to relate to others. A study by Reznikoff (1955) of women with cancer of the breast similarly revealed that a significantly higher proportion of cancerous women than controls had had a sibling die in childhood.

It was decided to test the hypothesis that the cancer patient has suffered a psycholigical blow in childhood more often than chance would allow by making a specific prediction from it. The prediction is that the birth of the next youngest sibling occurred earlier in the lives of patients with cancer than in those of controls. The reasoning is basically simple. The birth of a younger sibling, with the consequent perceived loss of parental energy and time, is a traumatic event. Other things being equal, the earlier this occurs, the more severe the trauma. The child has had less time to be the baby and is younger and therefore less able to cope with his emotions; the younger the child, the more his feelings are on a preverbal and magical level and, therefore, more difficult to deal with.

METHOD

It was decided to choose a group of families which had been carefully evaluated from a medical standpoint, and which had at least one cancerous individual among the offspring. A set of pedigrees was chosen from Jacobsen's (1946) extensive data.[2] He studied 200 cancer patients, treated in Copenhagen hospitals, on whom very complete family records were available. Certain of Jacobsen's families were excluded. These were families in which the cancerous individual was the youngest sibling, a twin, or an only child, and those where the ages of some of the siblings were missing.

The controls were noncancerous offspring in the same families. It was believed that using these subjects would minimize genetic factors and make any systematic impact of other carcinogens in the data quite improbable.

As an additional test of the hypothesis, data on the *S*s used in Reznikoff's study (1955) were reviewed with specific respect to the next youngest sibling. The *S*s in this research were 50 women attending the Breast Clinic of Memorial Center for Cancer and Allied Diseases, New York, for the first time. It was eventually found that 25 of these women had malignant neoplasms, while the remaining half were subsequently diagnosed as having benign disorders. Also included in this study were 25 women from the Strang Detection Center declared to be entirely free of any breast pathology. The three groups were approximately the same in their age distributions. There were no group differences in average number of siblings, although the cancer *S*s significantly less frequently had two or fewer sibs. The cancer and benign *S*s were of comparable economic status, having been screened to insure that their incomes met the eligibility requirements of the Breast Clinic.

RESULTS

A chi square test of the frequencies in Table 1 shows that there is a statistically significant difference in duration of being the youngest child between the cancerous individuals and the controls. Individuals with cancer tend to have a shorter period of being the youngest child than do their cancer-free siblings. It is to be noted, however, that the difference falls almost entirely within the first interval of 10–23 months, which encompasses only 27% of the cancer cases and 9% of the controls.

As can be seen from Table 2, a significant difference was also obtained between Reznikoff's three groups, compared for age of next youngest sibling, using an analysis of variance design. From Table 3, which contrasts these three groups considered two at a time, it is further revealed that the cancer *S*s were significantly closer in age to their next youngest sib than were the cancer-free individuals. Neither the comparison between the cancer and benign disorder groups nor between the benign disorder and cancer-free *S*s resulted in significant differences, although the former approaches significance.

---

[1] This paper was written under the auspices of a grant from the Ayer Foundation, Inc.

[2] The writers wish to express their appreciation to the staff of the library of the American Cancer Society for their courtesy and help in selecting this set of pedigrees.

TABLE 1

FREQUENCY OF INDIVIDUALS WITH AND WITHOUT CANCER AS RELATED TO THE NUMBER OF MONTHS BETWEEN OWN BIRTH AND THAT OF THE NEXT YOUNGEST SIBLING[a]

	Months Between Own Birth and That of Next Youngest Sibling												
	10–23		24–35		36–47		48–57		56–64		64+		Total
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	
Individuals with cancer	21	27	28	36	13	17	8	10	2	3	6	8	78
Cancer-free siblings	21	9	97	40	57	24	25	11	17	7	17	7	234

$\chi^2 = 18.32$, $df = 5$, $p < .01$.

[a] Jacobsen's families.

TABLE 2

COMPARISON OF CANCER, CANCER-FREE, AND BENIGN DISORDER GROUPS WITH RESPECT TO THE NUMBER OF MONTHS BETWEEN OWN BIRTH AND THAT OF THE NEXT YOUNGEST SIBLING

Source of Variation	*df*	Mean Square	*F*
Between groups	2	2558.51	3.78
Within groups	53	677.74	

$p < .05$.

TABLE 3

COMPARISON OF CANCER, CANCER-FREE, AND BENIGN DISORDER GROUPS CONSIDERED TWO AT A TIME WITH RESPECT TO THE NUMBER OF MONTHS BETWEEN OWN BIRTH AND THAT OF NEXT YOUNGEST SIBLING

Comparison	Mean difference months	*t*	*p*
Cancer with cancer-free	22.73	2.69	.02
Cancer with benign disorder	15.47	1.81	.10
Benign disorder with cancer-free	7.26	.85	.40

Note.—Means: cancer 26.53 months; cancer-free 49.26 months; and benign disorder 42.00 months.

DISCUSSION

A growing body of literature suggests that there are some correlations between personality organization and the occurrence of malignant disease. Two recent reviews (Kowal, 1955; LeShan, 1959) show that a large diversity of studies, using different methods, have come to this conclusion. Some studies seem to show indirectly that an early emotional trauma is found in a significant percentage of individuals who later develop cancer. Evans (1926) states that her cancer patients, who were in Jungian depth analysis, often regressed to a primary identification with the mother. In breast cancer patients, Bacon, Rennecker, and Cutler (1952) found a great deal of unresolved tension with the mother-figure. Cobb (1952) stated that her cancer patients regarded emotional relationships as dangerous and showed an absence of cathectic attachments. The papers of LeShan and Worthington (1956) and of Reznikoff (1955) point directly to an early traumatic event as being correlated with neoplastic disease.

It should be made clear that there is no implication here of a single cause of cancer. Our data suggest that under certain conditions, early emotional trauma may interact with other factors and play a part in the etiology of a later malignancy. The possible meanings of psychological events in the development of cancer have been partially explored by Meerloo (1954) and others, but much further research is needed to expand our gradually increasing awareness that psychological factors apparently must be viewed as being of some significance in the development of malignant disease, and to integrate this knowledge with the physiological findings in this field.

REFERENCES

BACON, C. L., RENNECKER, R., CUTLER, M. A psychosomatic survey of cancer of the breast. *Psychosomat. Med.*, 1952, **14,** 543–560.

COBB, B. A social-psychological study of the cancer patient. Unpublished doctoral dissertation. Univer. of Texas, 1952.

EVANS, E. *A psychological study of cancer.* New York: Dodd, Mead, 1926.

JACOBSEN, O. *Heredity in breast cancer.* London: Lewis, 1946.

KOWAL, S. Emotions as a cause of cancer. *Psychoanal. Rev.*, 1955, **42,** 217–227.

LESHAN, L. Psychological status as factors in the development of neoplastic disease: A critical review. *J. Nat. Cancer Inst.*, 1959, **22,** 1–18.

LESHAN, L., & WORTHINGTON, R. Some recurrent life history patterns observed in patients with malignant disease. *J. Nerv. Ment. Dis.*, 1956, **124,** 460–465.

MEERLOO, J. Psychological implications of malignant growth. *Brit. J. med. Psychol.*, 1954, **27,** 210–215.

REZNIKOFF, M. Psychological factors in breast cancer. *Psychosom. Med.*, 1955, **17,** 96–108.

(Received January 9, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SOME RELATIONSHIPS BETWEEN GSR CONDITIONING AND JUDGMENTS OF ANXIETY[1]

HAROLD GILBERSTADT AND GENE DAVENPORT

University of Minnesota and Veterans Administration Hospital, Minneapolis

PREVIOUS studies have demonstrated differences in conditionability between samples of individuals grouped on the basis of anxiety. Classical conditioning procedures have been used, employing the eyelid reflex (Spence & Farber, 1953; Spence & Taylor, 1953; Taylor, 1951; Taylor & Spence, 1954) as well as the galvanic skin response (GSR) (Bitterman & Holtzman, 1952; Howe, 1958; Lykken, 1957; Welch & Kubis, 1957). The samples have included both normals and psychiatric patients. Anxiety has been assessed by self-inventories such as the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953) and by behavior rating scales such as that of Buss, Wiener, Durkee, & Baer (1955).

The mechanics of conditioning procedures have varied. However, two publications (Moeller, 1954; White & Schlosberg, 1952) have presented evidence that the probability of successful conditioning is increased by using a one-half second interval between the onset of the conditioned (CS) and unconditioned stimulus (US).

The aim of the present study was to investigate further the relationship between GSR conditioning and various clinical judgments of anxiety in a sample of psychiatric patients conditioned under presumably optimal procedures.

METHOD

Sample. Nineteen out of 23 consecutive admissions to the open wards of the Minneapolis VA Hospital constituted the sample. The four patients who were not used included two who showed no GSR to shock or buzzer and two who had to be eliminated because of apparatus irregularities. The sample included three depressive reactions, three psychophysiologic reactions, seven anxiety reactions, two emotionally unstable personality reactions, and one each of paranoid state, adult situational reaction, and incipient schizophrenic reaction. Mean Shipley-Hartford IQ was 112 on the Sines (1958) norms, with a range of 96 to 126. Mean age was 32 with a range of 21 to 57.

Apparatus. The apparatus consisted of a Fels Dermohmeter 22A and an Esterline-Angus recorder. A 2 cm. electrode was centered on each palm, and the discharge from a 2.0 microfarad condenser charged from a 75 volt source was applied across the third phalanges of the second and fourth fingers of the dominant hand.

Conditioning. Instructions were given which represented the investigation as a routine hospital test of physiological functioning. The conditioning procedure followed Lykken (1955, 1957) with the exception that a one-half second CS-US interval was used in place of the five-second interval. Adaptation consisted of ten trials in which a buzzer (No. 1) and a different-toned buzzer (No. 2) were presented intermittently but not followed by shock. Conditioning consisted of shock reinforced presentations of buzzer No. 1, interspersed with four nonreinforced presentations of buzzer No. 2. Extinction consisted of 24 presentations of the two buzzers without shock.

Measurement. The principal measure employed was an adaptation of the conditioning score devised by Lykken (1955) and consisted of the difference between (*a*) the ratio of the average GSR to buzzer No. 1 on the first three presentations of that buzzer during extinction, divided by the average GSR to shock on the last three presentations of shock during conditioning and (*b*) the ratio of the average GSR to buzzer No. 1 on the last three adaptation trials, divided by the GSR to shock on the first three conditioning trials.

Assessments of clinical anxiety. First, three clinical psychologists without knowledge of conditioning results independently categorized patients as high, medium, or low anxiety types by inspection of psychology folders containing admission notes and routine admission tests, including the MMPI. No projective tests were available to the judges. Second, the Buss (1955) rating procedure was used by the psychiatrist in charge of the case who rated his patient on the basis of his brief admission interview. Finally, patients were sorted into three approximately equal groups on the basis of the MAS items from the MMPI.

RESULTS AND DISCUSSION

The relationships between the conditioning measure and the various anxiety measures are shown in Table 1. It can be seen that high, medium, and low anxiety groups as ranked by Clinician A, Clinician B, and the combined ratings of the three clinicians were significantly different in conditionability ($p < .05$) by the Kruskal-Wallis Analysis of Variance by Ranks test (Siegel, 1956). Neither

---

[1] We are indebted to J. J. Jenkins from whom we received helpful suggestions.

TABLE 1

RATED ANXIETY OF PATIENTS ORDERED ACCORDING TO CONDITIONING RANK

Item	Conditioning Rank																			Kruskal-Wallis Test	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	H	P
Anxiety Criterion:																					
Clinician A	H	H	H	H	H	H	H	L	M	M	M	L	M	L	M	L	M	L	M	12.60	$P < .01$
Clinician B	H	H	M	M	H	H	L	L	H	M	M	M	M	L	M	L	M	L	M	6.34	$.02 < P < .05$
Clinician C	H	H	M	H	H	H	M	L	H	M	L	M	L	L	M	L	M	L	H	4.78	$.05 < P < .10$
Combined Clinicians	H	H	M	H	H	H	M	L	H	M	M	M	M	L	M	L	M	L	M	8.77	$.01 < P < .02$
MAS	M	H	L	H	M	H	[a]	H	H	M	L	M	M	L	H	L	M	L	H	2.91	$.20 < P < .30$
Psychiatric Rating	H	M	L	L	M	M	H	L	H	L	L	L	L	M	M	M	H	L	M	.17	$.70 < P$
Diagnosis:[b]	N	N	P	N	N	N	P	N	N	P	N	C	N	N	C	N	N	N	N		

Note.—H = High anxiety rating; M = Medium anxiety rating; L = Low anxiety rating.

[a] No MMPI taken by this patient.

[b] N = neurotic; P = psychotic; C = character disorder.

the rankings by the MAS nor the psychiatric ratings were significantly related to GSR conditioning.

Spearman rank correlation coefficients (Siegel, 1956) between basal resistance and the conditioning measure ($r = -.22$) and between basal resistance and the average shock response ($r = -.04$) were not significant, nor were there differences in extinction between the various anxiety groups.

It has been noted (Cronbach & Meehl, 1955) that while studying the validity of clinical constructs, it is important also to scrutinize the criterion measures. The results suggest that ratings of clinical anxiety by experienced clinicians based on psychological tests and history data (including the MMPI) covary with conditioning of the GSR. There is a suggestion in the present investigation, as in other recent investigations (Goldberg, 1959), that studies employing judges should have some basis for inferring the expertness of the judges with the particular instruments. In the present instance, Clinician A, who was presumed to be most competent and most experienced with the MMPI was apparently most successful, and Clinician **C,** who was presumed to be least competent and least experienced with the MMPI, was apparently least successful (See Table 1).

Present findings are consistent with those of Bitterman and Holtzman (1952), in which judgments of anxiety by clinicians with several types of information available were significantly related to GSR conditioning, whereas the MAS was not.

The psychiatric ratings based on clinical observation from brief admission interviews were not significantly related to conditioning in the present instance, although in the Welch and Kubis study (1947) clear-cut relationships were obtained, possibly because the psychiatrists had greater familiarity with the patients.

One factor which could in part account for the lack of significant differences in extinction between groups was revealed during posttest interviews. Some patients stated that during the long series of buzzers in the extinction trials they "expected" a few shock stimulations. Cook and Harris (1937) have shown that such a factor can override the effects of conditioning and extinction.

SUMMARY

Using presumably optimum GSR conditioning procedures with a one-half second US-CS interval and an ongoing sample of 19 psychiatric patients rated for clinical anxiety, (*a*) anxiety groups ranked on the basis of admission psychological data by clinical psychologists were found to be significantly different in conditionability; (*b*) anxiety groups ranked either on the basis of the Taylor Manifest Anxiety Scale or brief psychiatric admission interviews were apparently not significantly different on conditioning; and (*c*) anxiety groups revealed no significant differences in extinction.

REFERENCES

BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47,** 615–623.

BUSS, A. H., WIENER, M., DURKEE, ANN, & BAER, M. The measurement of anxiety in clinical situations. *J. consult. Psychol.*, 1955, **19,** 125–129.

COOK, S. W., & HARRIS, R. E. Verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21,** 202–210.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52,** 281–302.

GOLDBERG, L. R. The effectiveness of clinicians' judgments: the diagnosis of organic brain damage from Bender-Gestalt test. *J. consult. Psychol.*, 1959, **23**, 25–33.

HOWE, E. S. GSR conditioning in anxiety states, normals, and chronic functional psychotic subjects. *J. abnorm. soc. Psychol.*, 1958, **56**, 183–189.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. Unpublished doctoral dissertation, Univer. Minnesota, 1955.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. *J. abnorm. soc. Psychol.*, 1957, **55**, 6–10.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

SINES, L. K. Intelligence test correlates of Shipley-Hartford performance. *J. clin. Psychol.*, 1958, **14**, 399–404.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.

TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A., & SPENCE, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–503.

WELCH, L., & KUBIS, J. Conditioned PGR (psychogalvanic response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.

WHITE, CARROLL T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

(Received October 13, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

VERBAL REINFORCEMENT IN SCHIZOPHRENIA[1]

EDWIN COHEN
Veterans Administration Hospital, Iowa City, Iowa

AND BERTRAM D. COHEN[2]
Lafayette Clinic, Detroit

THE common clinical observation of impaired "reality contact" in schizophrenia implies a deficit in the intensity or direction of responsiveness to social cues. Social cues, and specifically those involved in verbal interactions, have been shown to provide reinforcement effects in a wide variety of experimental settings (Salzinger, in press). If schizophrenic patients manifest faulty responsiveness to social cues, it may be predicted that the reinforcing function of listener reactions (i.e., verbal reinforcement) will be negligible for them as compared with nonschizophrenic control *S*s. The present experiment was designed to test this prediction. The particular technique used was one with which reinforcement susceptibility had been previously demonstrated for general medical patients (Cohen, Kalish, Thurston, & Cohen, 1954) and for psychiatric patients selected on the basis of anxiety level scores rather than diagnosis (Taffel, 1955).

METHOD

Apparatus. The materials consisted of 160 3×5 white index cards, on each of which was typed a common verb (used by Cohen et al., 1954). All verbs were in the past tense. Six pronouns—I, We, He, They, She, and You—were placed below each verb. The order of pronouns was randomized for the series of 160 cards.

*Subjects. S*s were 32 male hospitalized patients from the psychiatric service of the VA Hospital in Iowa City. All were capable of reading and free of CNS pathology. By unanimous agreement of the staff—two psychiatrists and two clinical psychologists—half of the *S*s had been diagnosed as schizophrenic, the other half as neurotic. Both groups were unselected as to subtype of illness and were comparable in age and education. Mean age of the schizophrenics was 32, while that of the neurotics was 32.4; mean education was 11.3 and 10.5 years for the schizophrenics and neurotics respectively. The groups varied significantly in the number of days they had been hospitalized; schizophrenics mean time was 130 days, with a range of 10 to 523, whereas the neurotics were hospitalized for a mean of 57 days with a range of 18 to 166 days.

---

[1] Portions of the present paper were read at the Midwestern Psychological Association meetings in St. Louis, Missouri, May, 1956.

[2] At VA Hospital, Iowa City, Iowa, when study was conducted.

Procedure. Each *S* was informed that he was participating in a research project. He was then shown a sample card and instructed to make up a sentence containing the verb on the card and beginning with any one of the pronouns. *S* then constructed sentences in this fashion until the 160 cards were completed. *E* recorded each sentence verbatim. The cards were shuffled for each *S* to randomize the effect of order.

The 16 *S*s in each diagnostic category were randomly divided into two equal groups. One group in each diagnostic category was reinforced for any sentence begun with I or We, and the other group for similar use of He or They. During the acquisition phase (the first 80 cards), *E* said "Good" in a flat, unemotional tone following any sentence which *S* began with the pronouns chosen to be reinforced. In order to determine a base level without reinforcement, however, reinforcement was not administered until after the first 20 card presentations.

The last 80 presentations constituted the extinction phase of the experiment. During this phase, reinforcement was discontinued for all *S*s. In order to test *S*'s awareness of the relation between his responses and those of *E*, each *S* was asked at the completion of the series which pronoun he had used most, as well as specific questions about his ideas concerning the purpose of the experiment and *E*'s responses.

TABLE 1

SUMMARY OF ANALYSIS OF ACQUISITION DATA

Source of Variation	*df*	*ms*	*F*
Between Subjects	31		
Diagnosis	1	82.89	
Methods	1	2.82	
Diagnosis × Methods	1	.38	
Error (b)	28	34.89	
Within Subjects	96		
Trials	3	32.15	3.96*
Trials × Diagnosis	3	28.57	3.52*
Trials × Method	3	5.38	
Trials × Diagnosis × Method	3	7.70	
Error (*w*)	84	8.12	
Total	127		

* Significant 2% level.

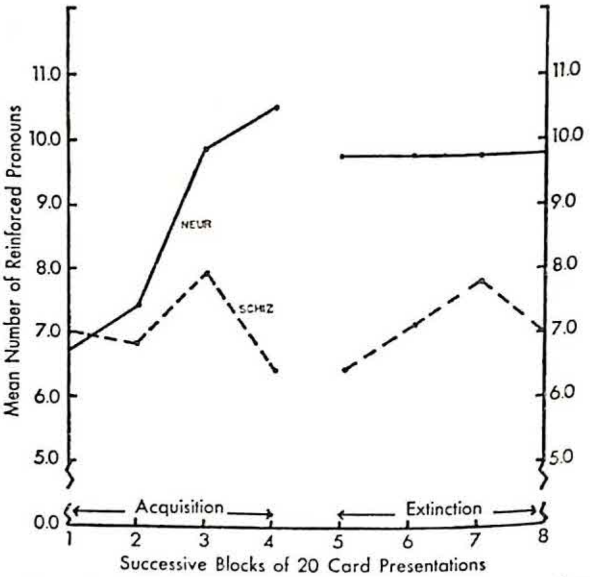


FIG. 1. Mean numbers of reinforced pronouns for successive blocks of 20 card presentations.

RESULTS

In the treatment of the data, the number of reinforced pronouns occurring within 8 successive blocks of 20 cards each were used as the criterion measures, taking the first four blocks as the acquisition and the last four as the extinction phases. A summary of the analysis of the acquisition trials is shown in Table 1. Neither Methods nor any of its interactions with other factors approaches reliable significance. It may therefore be safely assumed that the *S*s conditioned equally to both classes of pronouns, I-We and He-They. The *F* values for Trials and Trials-by-Diagnosis are each significant at the 2% level. It seems, therefore, that conditioning did occur, but it occurred differentially for the two diagnostic groups. Since it appeared from inspection of Fig. 1 that the curve for the schizophrenic *S*s was not ascending, the data for this group alone were further analyzed. The two methods of conditioning were combined, since the initial analysis had shown no differences attributable to methods. The *F* value for Trials was clearly not significant, allowing the inference that no reliable conditioning occurred for the schizophrenic group. It may then be concluded, because the over-all Trials effect for all groups is significant, that conditioning occurred in the neurotic group.

Since the neurotic group had demonstrated acquisition, their data for Trial Blocks 4 through 8 were analyzed for extinction. As suggested in Fig. 1, no reliably significant *F* was obtained for Trials. No evidence of extinction was found.

Initial questioning of *S*s with regard to the purpose of the experiment and the possibility that *E* had influenced their responses in any way revealed no clear evidence of awareness of the reinforcement contingency. Further questioning of a more specific nature, in which *E* explicitly asked if his saying "Good" had any effect on *S*'s responses, did result in "insight" statements by one schizophrenic and five neurotics. The acquisition scores for these *S*s inicated no better conditioning than the scores for the remaining 26 who furnished no evidence of awareness.

DISCUSSION

The acquisition results confirm the hypothesis that reinforcement effects are generally negligible for schizophrenic patients. This held true for both response classes to which the reinforcer was applied (either first or third person pronouns). In contrast, the neurotic groups showed significant acquisition curves. Again, the response class made no difference.

The combined acquisition curve for the two neurotic groups resembles that obtained for a group of nonpsychiatric patients reported in a previous paper (Cohen et al., 1954). The combined acquisition curve for the schizophrenic groups resembles that obtained for the nonreinforcement control group in the previous study. In other words, saying "Good" to the schizophrenic *S*s was apparently functionally equivalent to saying nothing to nonschizophrenic *S*s.

The failure to demonstrate extinction as a consequence of nonreinforcement, as seen in the results for the schizophrenic group, is to have been expected since no acquisition was demonstrated. On the other hand, acquisition was demonstrated by the neurotics, and yet extinction was not achieved under nonreinforcement. This finding is in accord with data for nonpsychiatric patients under a comparable extinction procedure (Cohen et al., 1954). The possible explanation cited in that study, referring to the existence of continuing sources of reinforcement inherent in the procedure, also applies to this study. Another possible explanation involves the relatively low frequency of the reinforced response at the initiation of the extinction procedure. It may be noted from Fig. 1 that the frequency of I and We for the neurotic *S*s was less than 11 for 20 cards at the time extinction was initiated. Since a previous study (Cohen et al., 1954) continued the reinforcement procedure for another 80 cards without any increase in frequency of the reinforced pronoun class, the solution to increasing its frequency would not seem to lie in additional reinforcement trials prior to initiation of extinction procedures. One possible solution may be to decrease the number of possible different responses, thus raising the initial level and the final acquisition level as well. For example, You and She could be eliminated. and I, We, He, and They retained.

Certain limits on the generality of the present results, regarding the nonsusceptibility to reinforcement of schizophrenic patients, should be specified. These include the limits set by (*a*) properties of the stimulus class employed as a reinforcer and (*b*) properties of the response class subjected to reinforcement operations.

In the present experiment, the reinforcing stimulus (Good) was one that connotes social approval. *E*'s explicit responses were administered so as to provide reward for the desired response class rather than punishment of responses in a competing class. Results summarized by Rodnick and Garmezy (1957) suggest that schizophrenic *S*s are especially sensitive to verbal punishment. It is possible that the aversive properties of verbal punishment could result in more normal acquisition curves for schizophrenic *S*s. A recent study by Leventhal (1958) provides data in support of this prediction.

Studies of motor learning (Cohen, 1956), reaction time (Rosenbaum, Mackavey, & Grisell, 1957), and operant behavior (Lindsley, 1956) of schizophrenic *S*s support the hypothesis of Hunt and Cofer (1944) that psychological deficit in schizophrenia reflects a loss of social motivation based on a process of extinction of responsiveness to acquired incentives and rewards. If schizophrenic patients have not incurred some irreversible impairment of the ability to respond to acquired reinforcers, one might predict that the experimental pairing of a stimulus like the word Good with a primary reinforcer would result in its acquisition (or reaquisition?) of reinforcing properties.

In an interview-type setting, Salzinger and Pisoni (in press) applied verbal reinforcing stimuli (Recognition and Agreement) to schizophrenic patients' verbal responses of a "personal-affective" type. Significant operant acquisition effects were demonstrated. No changes were found in nonaffective self-reference responses during the same interview. In the present experimental procedure, the reinforcer was applied to responses in the prearranged classes, regardless of any specific affective meanings the responses may have had for the individual patients.

A study by Hartman (1955) suggests that the initial (operant level) response subjected to reinforcement operations may be of particular importance for schizophrenic patients. Using a technique similar to that of the present study, Hartman subdivided his *S*s in accordance with the initial levels of preference they showed for the to-be-reinforced pronoun class. For his schizophrenic *S*s, he found significant acquisition effects for the subgroup which showed a high initial preference for the reinforced pronoun class. In the present study, reinforcement was administered to responses without regard for prereinforcement preferences.

Impaired susceptibility to verbal reinforcement, in view of the present results, warrants consideration as an objective, experimentally manipulable correlate of schizophrenic behavior. It is recognized that many investigators conceive of schizophrenic impairment of responsiveness to social cues (e.g., disturbances in reality contact) in terms of changes in perceptual and cognitive variables. Studies can be designed to clarify the relations of these variables to those of verbal reinforcement. In any event, further systematic investigation of the various parameters of verbal reinforcement in normal and pathological groups appears both warranted and promising.

REFERENCES

COHEN, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 186–190.

COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.

HARTMAN, C. H. Verbal behavior of schizophrenic and normal subjects as a function of types of social reinforcement. Unpublished doctoral dissertation, State Univer. Iowa, 1955.

HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.) *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 971–1032.

LEVENTHAL, A. The effects of differential verbal reinforcement on psychiatric and nonpsychiatric hospital patients. Unpublished doctoral dissertation, State Univer. Iowa, 1958.

LINDSLEY, O. R. Operant conditioning methods applied to research in chronic schizophrenia. *Psychiat. res. Rep.*, 1956, No. 5, 118–139.

RODNICK, E., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones, (Ed.) *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1957. Pp. 109–184.

ROSENBAUM, G., MACKAVEY, W. R., & GRISELL, J. L. Effects of biological and social motivation in schizophrenic reaction time. *J. abnorm. soc. Psychol.*, 1957, **53**, 364–368.

SALZINGER, K. Experimental manipulation of verbal behavior: a review. *J. gen. Psychol.*, in press.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, in press.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

(Received April 2, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

EGO DISJUNCTION:

A FAILURE TO REPLICATE TREHUB'S RESULTS

WILBERT E. FORDYCE
Veterans Administration Hospital, Seattle, Washington

AND WILLIAM R. CROW
University of Washington

TREHUB (in press) introduced the concept of ego disjunction (ED), defining it as "the concurrence in an individual of two or more relatively strong, relatively long enduring needs or characterological dispositions with mutually incompatible objectives." ED was measured by the Edwards Personal Preference Schedule (EPPS) Edwards, 1954), on which seven pairs of a prior incompatible needs were taken to indicate the level of ED. Trehub found a monotonic linear relationship between ED scores on the EPPS and diagnostic category, ranging from low ED for college students, up through adolescents, neurotics, character disorders, and schizophrenics. The last three groups were patients at a VA Neuropsychiatric hospital.

This study replicated that part of Trehub's study involving patient groups. In addition to the ED measure, a series of quantitative indices of disorganization based on the MMPI were compared with ED scores and with the diagnostic criterion. MMPI measures used (all expressed in *T* scores) were Pt-Sc plus 100 (Delta); Sc plus 2 (Sc-Pt), with Sc scores above 90 considered as 90

TABLE 1

MEANS AND STANDARD DEVIATIONS OF DIAGNOSTIC GROUPS ON ED AND SELECTED MMPI SCORES

Groups		N	Scores				Social Desirability
			Ego Disjunction	Delta	Gamma	Triad	
Neurotic	Mean	22	52.59	98.91	73.68	202.68	26.86
	SD		23.82	7.02	17.69	36.24	7.38
Character Disorder	Mean	31	50.42	100.16	68.06	193.48	27.94
	SD		22.93	11.66	30.16	34.59	6.65
Acute Schizophrenia	Mean	20	51.85	96.60	84.90	223.35	23.05
	SD		20.61	8.14	22.82	38.17	7.50
Chronic Schizophrenia	Mean	31	59.48	86.94	101.19	218.94	25.03
	SD		28.57	13.44	34.95	43.89	7.58

(Gamma); Pa plus Pt plus Sc (Triad); and the 39-item form of the Edwards (1953) Social Desirability scale (SD).[1]

Initially the diagnostic groups studied were drawn from a population of adult, male, veteran, neuropsychiatric patients at the VA hospital (GM and S), Seattle, Washington. There were 20 acute schizophrenics (Ac-Sc), 31 character disorders (Ch-Dis), and 22 neurotics (Neu) as classified by a single, final diagnosis assigned by the hospital staff. No cases with multiple psychiatric diagnoses were used. Later, a further sampling of more chronic schizophrenic patients was made to extend the range of severity of disorganization tapped by the tests. Under the same criterion of final hospital diagnosis, chronic male schizophrenics (Ch-Sc) were added—21 from Western State Hospital, Steilacoom, Washington, and 10 from the VA hospital (NP), American Lake, Washington.[2] These two subsamples were combined.

All subjects performed on the EPPS and the booklet form of the MMPI either the same day or within a very few days of each other.

RESULTS

The data failed to duplicate Trehub's findings. Table 1 shows the means and standard deviations on ED and the MMPI scores for each diagnostic group. Table 2 shows the *t* test comparisons of the means, indicating no differences between any of the criterion groups on the ED measure. In contrast, the MMPI indices of disorganization more consistently discriminate the schizophrenic groups from the character disorder group, though not from the neurotic group.

DISCUSSION

The failure of the ED measure to discriminate schizophrenic patients from the character disorder or neurotic patient groups raises a serious question as to the validity of the EPPS measure of the ego disjunction construct. While no adolescent or college student groups were included in this study, the range of presumed levels of ego disjunction represented in the four patient groups studied should be considerable. The success of the MMPI indices in discriminating the groups lends support to the inference that a range of sickness levels or ego disjunction does exist in the groups studied. There was no criterion contamination in the chronic schizophrenic group because the MMPI was administered after diagnoses were made. In the other groups, the MMPI as a part of the diagnostic workup process may have contributed in varying amounts to final diagnosis. Criterion contamination could account for differences found between neurotic and character disorder groups and the acute schizophrenic group. It is obvious, however, that far more than the MMPI was involved in arriving at a final diagnosis.

Failure of ED to discriminate when the MMPI indices succeeded suggests that sampling differences may have influenced Trehub's original findings. Differences in diagnostic criteria may have entered into these conflicting findings. It seems possible that cultural differences in test behavior between patients from New England and from the Pacific Northwest may contribute to the different findings of the two studies.

The fair success of the MMPI measures used here in discriminating psychotics from nonpsychotics offers some promise as a quantitative, objective criterion for use in other studies.

---

[1] The Delta and Gamma scores were defined arbitrarily and are designed to treat differently the shape and elevation factors of the Pt and Sc scales.

[2] We are indebted to F. E. Shovlain, Superintendent; R. W. Brown, Clinical Director; and Mack F. Knutsen, Supervising Psychologist, all of Western State Hospital; and to Leon S. Diamond, Chief of Professional Services and John Marks, Chief, Psychiatric Research Unit, VA hospital, American Lake, Washington, for their help in providing data for this study.

TABLE 2

DISCRIMINATION OF DIAGNOSTIC GROUPS BY t TESTS ON ED AND MMPI INDICES

Groups	Scores								*SD*	
	ED		Delta		Gamma		Triad			
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Neu–Ch-Dis	.384	.80	.440	.70[a]	.769	.50[a]	.899	.40	.545	.60
Neu–Ac-Sc	.104	.90	.962	.40	1.475	.10	1.756	.10	1.614	.20
Neu–Ch-Sc	.907	.40	3.750	.01[a]	2.650	.02[a]	1.399	.20	.855	.40
Ch-Dis–Ac-Sc	.222	.90	1.167	.30	2.092	.05	2.834	.01	1.224	.30
Ch-Dis–Ch-Sc	1.354	.20	4.080	.01	3.935	.01	2.496	.02	1.582	.20
Ac-Sc–Ch-Sc	1.013	.40	2.833	.02[a]	1.810	.10[a]	.361	.80	.896	.40

[a] Correction for unequal variances made by calculating df as $N - 1$ of the smaller group.

SUMMARY

Trehub's EEP-derived ego disjunction measure was compared with a series of MMPI-derived measures of ego disorganization on groups classified by psychiatric diagnosis. In contrast to Trehub's original study, the ego disjunction measure failed to discriminate among neurotics, character disorders, acute schizophrenics, and chronic schizophrenics. All MMPI measures except the Social Desirability scale discriminated schizophrenics from nonschizophrenics.

It was concluded that sampling differences, differences in diagnostic practice, or cultural differences in patient groups studied could account for the conflicting findings of this study and the original work by Trehub.

Success of the MMPI indices in discriminating psychotics from nonpsychotics offers promise of the utility of these indices in other studies.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

EDWARDS, A. L. *Personal preference schedule.* New York: Psychological Corp., 1954.

EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

TREHUB, A. Ego disjunction and psychopathology. *J. abnorm. soc. Psychol.*, in press.

(Received April 20, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

A NOTE CONCERNING HOMOGENEITY OF MEMBERSHIP AND GROUP PROBLEM SOLVING

MARVIN E. SHAW

Massachusetts Institute of Technology

HOMOGENEITY of group members with regard to individual differences has long been recognized as an important variable in group behavior (Hemphill, 1950). In a recent article, Hoffman (1959) reports results supporting the hypothesis that homogeneity results in less efficient problem solving than does nonhomogeneity of membership. Citing the findings of Pelz (1956) to support his view, he suggests that "... the results reported in the present study are probably generalizable well beyond the limited population of college students who supplied the data." (Hoffman, 1958, p. 31). It was this statement which motivated the writer to report some of his own observations of relations between homogeneity and group effectiveness.[1]

During the past few years, the writer has conducted a series of experiments attempting to study the joint effects of group structure and

---

[1] The purpose of this note is not to criticize Hoffman's excellent report, but only to offer a word of caution regarding the generalizability of the results and a word of encouragement to others who may be interested in this problem.

TABLE 1

THE RELATION OF MEMBER HOMOGENEITY TO GROUP PERFORMANCE AND SATISFACTION

Homogeneity Variables	Structural Conditions	N	Group Process Variables		
			Trials to Target	Time Scores	Ratings of Satisfaction
Intelligence (Shaw, 1959a)	Power:				
	Centralized	20	+.07	+.38	−.49*
	Decentralized	20	−.07	+.08	+.21
	Communication:				
	Centralized	24	—	.00	+.21
	Decentralized	24	—	+.26	+.30
Acceptance of Authority (Shaw, 1959a)	Power:				
	Centralized	20	−.11	−.17	−.08
	Decentralized	20	+.48*	+.53**	−.59**
	Communication:				
	Centralized	24	—	.00	+.02
	Decentralized	24	—	+.27	−.11
Individual Prominence (Shaw, 1959b)	Power:				
	Centralized	22	−.22	+.06	−.18
	Decentralized	22	−.08	−.26	+.20

Note.—Signs of correlations have been adjusted so that a positive correlation indicates increasing efficiency or satisfaction with increasing homogeneity.

* $p < .05$.

** $p < .01$.

composition upon group efficiency in problem solving situations. (Shaw, 1959a, 1959b). While not related to the primary purpose of the investigations, the data were analyzed systematically to determine the relationship between various estimates of homogeneity and problem solving effectiveness under varying structural conditions. The results of these analyses are given in Table 1. The measure of homogeneity used in computing the correlations reported was the average deviation of scores earned by members of the group on measures of intelligence (the verbal scale of the Scholastic Aptitude Tests), acceptance of authority, and individual prominence. Essentially the same results were obtained, however, when other indices of homogeneity were used, such as range and standard deviation. Measures of group process variables were number of trials to criterion, total time required by the group to complete the assigned task, and average ratings of satisfaction by group members.

A number of comments may be made in connection with the results reported in Table 1. First, there seems to be no systematic relation between homogeneity and group efficiency and satisfaction. Degree and direction of correlations vary with the group structure and the particular characteristic under consideration.

Second, most of the correlations fail to meet the usual requirements for statistical significance. This may be due to the small number of cases in each sample, to the fact that the homogeneity estimate is based upon a single characteristic rather than a profile as were Hoffman's measures, or to the fact that no relation exists. It is interesting to note, however, that the one condition yielding uniformly significant correlations (acceptance of authority in a decentralized power structure) is precisely the one where such a relation might be expected on a priori grounds.

Third, centralized structures appear consistently to produce correlations that differ from those found in decentralized structures, although these differences are mostly unreliable.

These findings point to the difficulties in generalizing the results of a single study or even several studies beyond the limits imposed by the experimental situations. It may be, of course, that the effects of profile homogeneity will prove to be unchanged by such extraneous variables as group structure, group task, and the like; however, it seems far more likely that the effects of homogeneity, like so many other variables in group behavior, will be found to vary with the particular conditions under which its effects are measured.

The general problem of homogeneity as a variable in group behavior is a tremendously important one and deserves extensive investigation. The limited evidence reported here, along with the findings of Hoffman (1959) and Pelz (1956), indicates that the results of such studies will richly repay the investigator for his efforts.

REFERENCES

HEMPHILL, J. K. The measurement of group dimensions. *J. Psychol.*, 1950, **29**, 325–342.

HOFFMAN, L. R. Homogeneity of member personality and its effect on group problem-solving. *J. abnorm. soc. Psychol.*, 1959, **58**, 27–32.

PELZ, D. C. Some social factors related to performance in a research organization. *Administrative Sci. Quart.*, 1956, **1**, 310–325.

SHAW, M. E. Acceptance of authority, group structure and the effectiveness of small groups. *J. Pers.*, 1959, **27**, 196–210. (a)

SHAW, M. E. Some effects of individually prominent behavior upon group effectiveness and member satisfaction. *J. abnorm. soc. Psychol.*, 1959, in press. (b)

(Received March 4, 1959)

THE BODY IMAGE BARRIER CONCEPT[1]

CHARLES R. EIGENBRODE AND WILLIAM G. SHIPMAN

Staunton Clinic, School of Medicine, University of Pittsburgh

A GOOD deal of the recent research on the body image has been done by Fisher and Cleveland. Their initial study of rheumatoid arthritics (Cleveland & Fisher, 1954) led them to focus on the external limits of the body image, which may serve unconsciously as a defensive barrier to ward off threats from the outside and to hold disturbing, hostile impulses within. A scoring system for Rorschach protocols was derived which yields a Barrier score—an index of the degree of concern a subject has about his body limits. This measure was validated on a group of patients with internal and external disorders (Fisher & Cleveland, 1955), then applied with interesting findings to cancer patients (Fisher & Cleveland, 1956b), to style of life (Fisher & Cleveland, 1956a), and to physiological reactivity (Fisher & Cleveland, 1958; Fisher, 1959). It has been assumed in these papers that the measure represents a psychological factor that is not only a strong determinant of locus of disease but is related to central personality tendencies. The purpose of the present study was to determine how well the Barrier score differentiates between patients with internal and external psychosomatic illnesses.

METHOD

Sample. The Rorschach records used in this study were obtained from the case files of Staunton Clinic. The medical record of each patient was checked so that assignment to one of the experimental groups could be made. No patient was included in the sample who (*a*) had symptoms of both internal and external disease, (*b*) was below 16 or above 50 years of age, or (*c*) had less than 10 responses in his Rorschach. Cases were drawn so that the two patient groups were similar in age, education, and IQ. Two groups of records were obtained: 54 cases of psychosomatic skin disorders (26 atopic eczema, 16 chronic urticaria, and 12 miscellaneous skin disorders) and 29 cases of internal psychosomatic disorders (18 stomach ulcer, 8 ulcerative colitis, and 3 genitourinary disease). Emotional factors were considered by the physicians to be the major factor in all skin cases.

Scoring. After learning the scoring system, the authors collaboratively scored the 83 records, which had been randomly mixed and from which all primary identification material had been removed. The number of scored responses was equated for the two groups by setting a maximum of 20 scored responses per record. This control was successful in that the median number of responses per record in both groups was 20. After three to five days, 30 of the 83 records were pulled at random and rescored as a check on the consistency with which the scoring system had been applied. This comparison yielded a Pearson correlation of .95. At this point, Fisher was consulted. His scoring of 10 of these records yielded a Pearson r of .84 with that of the authors. This consultation indicated that several basic changes in the scoring system used by the authors needed to be made in order to coincide more closely with Fisher and Cleveland's actual practice. This prompted a rescoring of the 83 records by the two authors individually. The extent of their agreement was indicated by a Pearson r of .80, which compares favorably with an r of .82 reported in the original monograph (Fisher & Cleveland, 1955).

---

[1] The authors wish to thank Seymour Fisher and Stanley King for personally clarifying the scoring system. Our associate, Leslie Berger, made many helpful suggestions in the latter stages of the project.

RESULTS AND DISCUSSION

A nonparametric median test was applied to the Barrier score data of both of the writers for the interval vs. the external disease groups. Neither of the obtained chi squares approached significance. Indeed, the obtained frequencies were almost identical with those expected by chance. In the light of this absence of difference in the Barrier scores obtained by patients with internal and external diseases, it seems doubtful that a concern with firmness of body boundaries plays a significant role in determining the location of an illness in the body.

Where does the reason for this discrepant finding lie: in the sample, in the measuring instrument, in the theory, or in some combination?

The sample differs from that used in the original study mainly in disease distribution and sex ratio. In this study, the sex ratio for the various diseases was kept similar to the known distribution of the disorders in the general population. But sex differences should not be an important source of discrepancy since both Fisher and Cleveland and the present investigators found no sex difference in the Barrier score.

The scoring system may have been misapplied or have intrinsic weaknesses that allow different results in different studies. The correlation of .84 between the initial scoring of the writers and Fisher, plus the subsequent rescoring, suggests that reliable scoring was not a problem. This turns the focus on the Barrier score itself, which seems open to considerable criticism. The scoring rules, *in detail*, have not been published. No retest reliability has been reported. And most important of all, no cross-validation of the scale seems to have been published by other investigators. The small size of the modal Barrier score, 4, and the range, ~~0 to 13~~ in this study, allows great instability of scores. Thus, a subject's classification as "high" or "low" may also be unstable.

This highlights another variable, the examiner. There is a great range in the amount of inquiry that is done during the administration of a Rorschach, and the presence or absence of additional descriptive material most certainly affects the opportunity for Barrier scores. To check on Barrier score differences between examiners, a nonparametric analysis of variance by ranks was done on the scores obtained by the seven examiners who tested 78% of the cases used here. The resultant Kruskal-Wallis H was statistically nonsignificant, and it was concluded, therefore, that differences between examiners did not cause significant differences in the number of Barrier responses given by the subjects in this sample.

Conceivably, the scale does not measure the Barrier concept very well. In many of the major scoring categories there is a great deal of apparent arbitrariness about which Rorschach responses fitting the category may be scored. For example, when considering animals with unusual skins, a mink is scored but a possum is not. Other inconsistencies emerge in the "unusual skin" category which may contribute a great deal to the total. And why is not the opposite end of the dimension scored? One would think that minus scores for such soft surfaced things as sponges and wool would improve the scoring. Neglecting to score responses at one end of the continuum should hypothetically greatly reduce the sensitivity of the measure. These various points lead one to conclude that there are some striking gaps in the application of the Barrier principle to the scoring system.

The theory itself warrants careful inspection. The scoring instructions refer to "an individual's perception of his body boundaries as *definite* and as possessing *protective* properties." There are obviously two features involved here, and it is a question whether these two features, definiteness and protectiveness, are necessarily related. A person may be well aware of his limits but not necessarily fearful of attack and defensively oriented. Presumably, body limits are clearer with emotional maturity, whereas a preoccupation with defense to external threat occurs at a less mature level. Similarly, the Barrier theory scores with equal weight, *containers* and *defensive covering*. Though these both involve a hard, encircling surface, the points of view involved seem quite different. One wonders if these two concepts really represent only one psychological variable or whether there are at least two factors involved. The point, then, is that the Barrier concept may entail different and not entirely consistent dimensions. This complexity of the over-all measure would be acceptable if the role of the subfactors were known and the measure correlated successfully with other variables. But under the present circumstances, some of the subcategories may be measuring the Barrier tendency while others, which do not, cancel the relation when the subcategories are combined.

SUMMARY

A check was made on whether the Rorschach Barrier scores of 54 patients with external psychosomatic disorders were greater than those of 29 patients with internal psychosomatic disorders. Statistical analysis showed no evidence of difference in the distribution of Barrier scores in the two groups. The potentialities of the Barrier approach cannot be truly assessed until ambiguities in the theory are clarified and until the scale is improved as a measuring instrument.

REFERENCES

Cleveland, S. E., & Fisher, S. Behavior and unconscious fantasies of patients with rheumatoid arthritis. *Psychosom. Med.*, 1954, **16**, 327–333.

Fisher, S. Prediction of body exterior vs. body interior reactivity from a body image schema. *J. Pers.*, 1959, **27**, 56–62.

Fisher, S., & Cleveland, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.*, 1955, **69** (17, Whole No. 402).

Fisher, S., & Cleveland, S. E. Body image boundaries and style of life. *J. abnorm. soc. Psychol.*, 1956, **52**, 373–379. (a)

Fisher, S., & Cleveland, S. E. Relationship of body image to site of cancer. *Psychosom. Med.*, 1956, **18**, 304–309. (b)

Fisher, S., & Cleveland, S. E. *Body image and personality.* Princeton, N. J.: Van Nostrand, 1958.

(Received August 26, 1958)